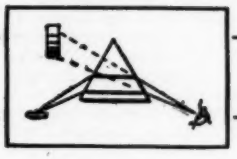
Journal of the OPTICAL SOCIETY *of* AMERICA

VOLUME 38
No. 8



AUGUST
1948

New Calculus for the Treatment of Optical Systems. VII. Properties of the N-Matrices . . . . . R. CLARK JONES 671

Multipupil in Phase Microscopy . . . . . HAROLD OSTERBERG 685

Achromatic Ultraviolet Microscope Objective . . . . . L. V. FOSTER AND E. M. THIEL 689

Calibration of a Photo-Multiplier Photometer . . . . . J. R. DEVORE 692

I.C.I. Specifications of Difference Limens for Munsell Hue, Value, and Chroma . . . . . JOSEPHINE G. BRENNAN AND SIDNEY M. NEWHALL 696

Paraxial Differential Transfer Coefficients of a Lens System . . . . . MARC G. DREYFUS 703

Irradiance Distribution in a Lloyd Mirror Interference Pattern . . . . . ROBERT N. WOLFE AND FRED C. EISEN 706

Recommended Footcandle Levels for Prolonged Critical Seeing . . . . MATTHEW LUCKIESH 712

Infra-Red Pupillometer . . . . . M. KOOMEN, R. TOUSEY, AND H. A. KNOLL 719

Distortion of Stereoscopic Spatial Localization . . . . KENNETH N. OGLE AND PAUL BOEDER 723

Reduction of Sphero-Chromatic Aberration in Catadioptric Systems . . ROBERT E. STEPHENS 733

Installation of the Ohio Valley Section of O.S.A. 735

Image Error Theory for Finite Aperture and Field. I. The Image of a Point; Geometry of the Wave Surface . . . . . MAX HERZBERGER 736

Letters to the Editor:

Measurement of Fluorescence Spectra . . . . . S. SAMBURSKY AND G. WOLFSOHN 739
Ultraviolet Phase Microscopy . . . . . A. H. BENNETT, D. L. WOERNLEY, AND A. J. KAVANAGH 739
Test on the Process of Photographing the Trajectory of a Rapidly Moving Distant Light Source . . . . . R. E. NUSBAUM AND J. S. BEGGS 740

Published monthly
at Prince and Lemon Sts., Lancaster, Pa.
by the
AMERICAN INSTITUTE OF PHYSICS
Incorporated
57 East 55 Street, New York 22, New York
FOR AND UNDER THE EDITORIAL SUPERVISION OF THE
OPTICAL SOCIETY OF AMERICA

THE OPTICAL SOCIETY OF AMERICA

Officers of the Society

RUDOLF KINGSLAKE, *President*
Eastman Kodak Company
Rochester, New York

WILLIAM F. MEGGERS, *Vice President*
National Bureau of Standards
Washington 25, D. C.

ARTHUR C. HARDY, *Secretary*
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

W. W. GRAEPER, *Treasurer*
Bausch & Lomb Optical Company
Rochester 2, New York

The Board of Directors consists of the above officers; the Editor of the Journal; the Past President; and five directors-at-large.

G. R. HARRISON, *Editor and Past President*
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

Directors-at-Large

FRANK BENFORD
General Electric Company
Schenectady, New York

F. M. BISHOP
Eastman Kodak Company
Rochester 4, New York

WALLACE R. BRODE
Ohio State University
Columbus 10, Ohio

A. H. BENNETT
American Optical Company
Buffalo 11, New York

E. H. LAND
Polaroid Corporation
Cambridge, Massachusetts

Local Sections

STANLEY S. BALLARD, *Secretary for the Local Sections*

Rochester Section

JOHN H. MCLEOD, *President*
HARRY D. POLSTER, *Vice President*
JAMES R. BENFORD, *Secretary*
LUVILLE T. STEADMAN, *Treasurer*

Detroit Section

DAVID L. FRY, *Chairman*
KARL W. BEYER, *Vice Chairman*
DANIEL L. HARMON, *Secretary*
W. R. O'NEILL, *Treasurer*

Niagara Frontier Section

GALEN PORTER, *President*
LELAND E. HAVENS, *Secretary*
CLAIR M. BIRDSALL, *Treasurer*

Ohio Valley Section

M. E. MERCHANT, *Chairman*
J. F. WOODRUFF, *Chairman-elect*
C. S. MILLS, *Secretary-Treasurer*

Patrons

AMERICAN OPTICAL COMPANY — BAUSCH & LOMB OPTICAL COMPANY
EASTMAN KODAK COMPANY

Corporation Members

ALLIED CHEMICAL AND DYE CORPORATION, NATIONAL ANILINE DIVISION, Buffalo, New York
AMERICAN CYSTOSCOPE MAKERS, INC., New York, New York
AMERICAN OPTICAL COMPANY, Southbridge, Massachusetts
ARGUS, INCORPORATED, Ann Arbor, Michigan
BAUSCH & LOMB OPTICAL COMPANY, Rochester, New York
BELL & HOWELL COMPANY, Chicago, Illinois
BELL TELEPHONE LABORATORIES, INC., New York, New York
CENTURY LIGHTING, INC., New York, New York
CORNING GLASS WORKS, Corning, New York
R. R. DONNELLEY & SONS COMPANY, Chicago, Illinois
E. I. DUPONT DE NEMOURS & COMPANY, INC., Chemical Department, Experimental Station, Wilmington, Delaware
EASTMAN KODAK COMPANY, Rochester, New York
EDMUND SALVAGE COMPANY, Oaklyn-Audubon, New Jersey
FAIRCHILD AVIATION CORPORATION, Jamaica, New York
FARRAND OPTICAL COMPANY, INC., New York, New York
FISH-SCHURMAN CORPORATION, New York, New York
GAERTNER SCIENTIFIC CORPORATION, Chicago, Illinois
W. & L. E. GURLEY, Troy, New York
INTERCHEMICAL CORPORATION, New York, New York
KEUFFEL & ESSER COMPANY, Hoboken, New Jersey
KOLLSMAN INSTRUMENT DIVISION OF SQUARE D COMPANY, Elmhurst, New York
NATIONAL CARBON COMPANY, Cleveland, Ohio
NATIONAL TECHNICAL LABORATORIES, South Pasadena, California
PERKIN-ELMER CORPORATION, Glenbrook, Connecticut
POLAROID CORPORATION, Cambridge, Massachusetts
PORTMAN MACHINE TOOL COMPANY, New Rochelle, New York
SIMPSON OPTICAL MANUFACTURING COMPANY, Chicago, Illinois
SOCONY-VACUUM LABORATORIES, Paulsboro, New Jersey
ARTHUR H. THOMAS COMPANY, Philadelphia, Pennsylvania
UNIVIS LENS COMPANY, Dayton, Ohio
ZENITH OPTICAL COMPANY, Huntington, West Virginia

THE OPTICAL SOCIETY OF AMERICA was organized in 1916 "*to increase and diffuse the knowledge of optics in all its branches, pure and applied, to promote the mutual interests of investigators of optical problems, of designers, manufacturers and users of optical instruments and apparatus of all kinds and to encourage cooperation among them.*" It aims to serve the interests of all who are engaged in any branch of optics from fundamental research to the manufacture of optical goods.

MEMBERSHIP

Regular Members: "Any person who has contributed materially to the advancement of optics is eligible for regular membership."

Associate Members: "Any person who is interested in optics is eligible for associate membership."

Corporate Members: "Any corporation desirous of encouraging the work of the Society is eligible for corporation membership."

Patrons: The payment, within any one year, of $1000 or more toward the expenses of the Society by any corporation or individual entitles such corporation or individual to the rank of patron.

The Society cordially invites to membership all who are interested in any branch of optics, either in research, in instruction, in optical or illuminating engineering, in the manufacture and distribution of optical goods of all kinds, or in physiological and medical optics.

ANNUAL DUES

Regular members, $8.50; Associate members, $6.00; Corporate members, $50.00. These dues include subscription to the Society's official journal, the *Journal of the Optical Society of America.*

CHANGES OF ADDRESS

All changes of address of members of the Optical Society of America should be communicated to the Secretary, Professor Arthur C. Hardy, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Journal of the
OPTICAL SOCIETY
of AMERICA

VOLUME 38, NUMBER 8 AUGUST, 1948

A New Calculus for the Treatment of Optical Systems.
VII. Properties of the N-Matrices

R. CLARK JONES
Research Laboratory, Polaroid Corporation, Cambridge, Massachusetts

(Received January 2, 1948)

The preceding papers of this series have examined the properties of an optical calculus which represented each of the separate elements of an optical system by means of a single matrix **M**. This paper is concerned with the properties of matrices, denoted by **N**, which refer not to the complete element, but only to a given infinitesimal path length within the element.

If **M** is the matrix of the optical element up to the point z, where z is measured along the light path, then the **N**-matrix at the point z is defined by

$$\mathbf{N} \equiv (d\mathbf{M}/dz)\mathbf{M}^{-1}. \tag{A}$$

Thus one may write symbolically,

$$\mathbf{N} = d\log\mathbf{M}/dz, \tag{B}$$

and

$$\mathbf{M} = \mathbf{M}_0 \exp(\int \mathbf{N} dz). \tag{C}$$

A general introduction is contained in Part I. The definition and general properties of the **N**-matrices are treated in Part II. Part III contains a detailed discussion of the important special case in which the optical medium is homogeneous, so that **N** is independent of z; Part III contains in Eq. (3.26) the explicit relation which corresponds to the symbolic relation (C). Part IV describes a systematic method, based on the **N**-matrices, by which the optical properties of the system at each point may be described uniquely and quantitatively as a combination of a certain amount of linear birefringence, a certain amount of circular dichroism, etc.; the method of resolution is indicated in Table I. Part V treats the properties of the inhomogeneous crystal which is obtained by twisting a homogeneous crystal about an axis parallel to the light path.

1.0 INTRODUCTION

THE preceding papers of this series[1] have been concerned with the properties of a calculus for the treatment of optical systems which modify the state of polarization of the light passing through them.

The calculus is based on the fundamental physical fact that the two principal components of the electric vector $\mathcal{E}'$ of the light emerging from an optical element are linear functions of the components of the electric vector $\mathcal{E}$ of the light entering the optical element, subject to certain conditions discussed in Paper V. Thus one may write

$$\mathcal{E}' = \mathbf{M}\mathcal{E}, \tag{1.1}$$

where the electric vector $\mathcal{E}$ is defined by

$$\mathcal{E} \equiv \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{1.2}$$

and where the matrix **M** is defined by

$$\mathbf{M} \equiv \begin{pmatrix} m_1 & m_4 \\ m_3 & m_2 \end{pmatrix}. \tag{1.3}$$

X and Y are the two principal transverse (com-

[1] I.—J. Opt. Soc. Am. **31**, 488–493 (1941); II.—*ibid.*, **31**, 493–499 (1941); III.—*ibid.*, **31**, 500–503 (1941); IV.—*ibid.*, **32**, 486–493 (1942); V.—*ibid.*, **37**, 107–110 (1947); VI.—*ibid.*, **37**, 110–112 (1947).

plex) components of the light, and m_1, m_2, m_3, and m_4 are complex numbers which are functions only of the optical element in question, and do not depend on X or Y.

In order to contrast the procedure used previously with that to be presented in this paper, let it be emphasized that the calculus made no effort to describe the state of polarization at every point in the optical system, but described the electric vector only at a number of places in the system—namely, in the spaces which separated the units designated as the elements of the optical system. This limitation occurred naturally because the fundamental mode of description by use of the matrix operators assigned a single matrix to describe the over-all behavior of the optical element, and provided no description of the optical behavior within subdivisions of the optical element.

This paper is concerned with a more detailed approach which permits the determination of the electric vector at each point along the path of the light. The matrix operator $\mathbf{M}$ of the complete element is represented as a line integral of a matrix $\mathbf{N}$ which is determinate at each point along the path of the light.

For the sake of simplicity of description, the presentation in this paper will be restricted to the case in which the optical element has plane and parallel surfaces, and in which the element is homogeneous in every sufficiently thin, plane and parallel section which is parallel with the surfaces. The formulation will thus include the case of light which passes normally through a crystal which has been twisted (not necessarily uniformly) about the normal to its surfaces and which has been stretched (not necessarily uniformly) along the normal. It will be evident, however, that the procedure may easily be generalized to hold for arbitrarily deformed crystals, and for the more general types of optical elements discussed in Paper V, by the introduction of suitable curvilinear coordinates.

2.0 DEFINITION AND PROPERTIES OF THE N-MATRICES

2.1 Definition of the N-Matrix

The crystal is assumed to have plane and parallel surfaces, and to be two-dimensionally homogeneous within every plane which is parallel with the surfaces. The origin of coordinates is at the surface at which the light enters, and the positive z axis extends into the crystal.

Consider first the matrix $\mathbf{M}_{z,z'}$ which represents the optical properties of a thin layer of the crystal whose surfaces have the coordinates z and z'. One then has by definition

$$\mathcal{E}_{z'} = \mathbf{M}_{z,z'}\mathcal{E}_z. \tag{2.1}$$

The $\mathbf{N}$-matrix at the coordinate z is then defined by

$$\mathbf{N}_z \equiv \lim_{z'=z} \frac{\mathbf{M}_{z,z'} - \mathbf{1}}{z' - z}. \tag{2.2}$$

Now let $\mathbf{M}_z$ represent the matrix of that part of the crystal which lies between the surface at which the light enters and the parallel surface at the depth z:

$$\mathcal{E}_z = \mathbf{M}_z\mathcal{E}_0, \tag{2.3}$$

where $\mathcal{E}_0$ is the electric vector of the light which enters the crystal. The matrix $\mathbf{M}_{z,z'}$ may then be written

$$\mathbf{M}_{z,z'} = \mathbf{M}_{z'}\mathbf{M}_z^{-1}. \tag{2.4}$$

Substitution of Eq. (2.4) in Eq. (2.2) then yields

$$\mathbf{N}_z \equiv \lim_{z'=z} \frac{\mathbf{M}_{z'} - \mathbf{M}_z}{z' - z}\mathbf{M}_z^{-1} \tag{2.5}$$

$$= (d\mathbf{M}_z/dz)\mathbf{M}_z^{-1},$$

or

$$\mathbf{N} \equiv (d\mathbf{M}/dz)\mathbf{M}^{-1}, \tag{2.6}$$

where the subscript z has been omitted because it is no longer needed. It is to be understood that both $\mathbf{M}$ and $\mathbf{N}$ are functions of z.

Equation (2.6) should be considered as the general definition of the matrix $\mathbf{N}$.

By writing Eq. (2.6) in the form

$$d\mathbf{M}/dz = \mathbf{N}\mathbf{M}, \tag{2.7}$$

it is evident that $\mathbf{N}$ is the operator which determines $d\mathbf{M}/dz$ from $\mathbf{M}$. It will be shown in the next section that $\mathbf{N}$ is also the operator which determines $d\mathcal{E}/dz$ from $\mathcal{E}$.

In Paper I of this series, it was shown that crystals representing various distinctive types of optical behavior corresponded to specific forms of the $\mathbf{M}$-matrices, and the various forms were exhibited. In the same way, there is a correspondence between the type of optical behavior

and the form of the **N**-matrices. The form of the N-matrices for eight different (exhaustive and mutually exclusive) types of optical behavior is shown in Table I.

As a specific example, consider the matrix

$$\mathbf{M}=\mathbf{S}(\tfrac{1}{2}\alpha z^2);\quad \mathbf{M}^{-1}=\mathbf{S}(-\tfrac{1}{2}\alpha z^2), \tag{2.A}$$

where **S** is the rotation matrix defined by Eq. (2.15). Then by differentiation of Eq. (2.A) one finds

$$d\mathbf{M}/dz=\alpha z\mathbf{S}(\tfrac{1}{2}\alpha z^2)\mathbf{S}(\tfrac{1}{2}\pi) \tag{2.B}$$

and, finally, by use of Eqs. (2.6), (2.A), and (2.B) one finds:

$$\mathbf{N}=\begin{pmatrix}0 & -\alpha z\\ \alpha z & 0\end{pmatrix} \tag{2.C}$$

2.2 Differential Equation for $\mathcal{E}$

Another important property of the matrix **N** is that it is the operator which determines $d\mathcal{E}/dz$ from $\mathcal{E}$. The proof follows.

By differentiation, one finds from Eq. (2.3)

$$d\mathcal{E}/dz=(d\mathbf{M}/dz)\mathcal{E}_0. \tag{2.8}$$

By the use of Eq. (2.6), the last relation becomes

$$d\mathcal{E}/dz=\mathbf{N}\mathbf{M}\mathcal{E}_0, \tag{2.9}$$

and by the use of Eq. (2.3) again, one obtains finally

$$d\mathcal{E}/dz=\mathbf{N}\mathcal{E}. \tag{2.10}$$

Accordingly, the **N**-matrix may be considered as a generalized wave number, since the wave number, defined as k in the expression

$$A=A_0\exp(i(\omega t-kz)), \tag{2.11}$$

is the number by which the amplitude is multiplied if the amplitude is differentiated with respect to z:

$$dA/dz=-ikA. \tag{2.12}$$

As a specific example of the use of Eq. (2.10) consider the matrix:

$$\mathbf{M}=\mathbf{S}(\omega z). \tag{2.D}$$

Substitution of Eq. (2.D) in Eq. (2.6) then yields:

$$\mathbf{N}=\omega\mathbf{S}(\tfrac{1}{2}\pi). \tag{2.E}$$

if the expression (2.E) for **N** is then substituted in the vector equation (2.10), this equation becomes equivalent to the following two scalar equations:

$$\begin{aligned} dX/dz&=-\omega Y,\\ dY/dz&=\omega X. \end{aligned} \tag{2.F}$$

Integration of the Eqs. (2.F) subject to the boundary condition that **M** must be equal to the unit matrix when z is zero, then leads back to the expression (2.D) for **M**.

2.3 Rotation of the Optical Element

It will now be shown that the **N**-matrices are transformed upon rotation of the element, in just the same way as the **M**-matrices.

Let a rotation from the positive x axis toward the positive y axis be termed positive. Then upon rotation of the optical element through the positive angle ω, it was shown in Paper I (p. 190) that the matrix **M** is transformed by the relations

$$\mathbf{M}'=\mathbf{S}(\omega)\mathbf{M}\mathbf{S}(-\omega), \tag{2.13}$$

$$\mathbf{M}'^{-1}=\mathbf{S}(\omega)\mathbf{M}^{-1}\mathbf{S}(-\omega), \tag{2.14}$$

where $\mathbf{S}(\omega)$ is the rotation matrix

$$\mathbf{S}(\omega)\equiv\begin{pmatrix}\cos\omega & -\sin\omega\\ \sin\omega & \cos\omega\end{pmatrix}. \tag{2.15}$$

Upon differentiating Eq. (2.13) with respect to z:

$$d\mathbf{M}'/dz=\mathbf{S}(\omega)(d\mathbf{M}/dz)\mathbf{S}(-\omega), \tag{2.16}$$

and then substituting Eqs. (2.14) and (2.16) in Eq. (2.6), one finds directly

$$\mathbf{N}'=(d\mathbf{M}'/dz)\mathbf{M}'^{-1}=\mathbf{S}(\omega)\mathbf{N}\mathbf{S}(-\omega), \tag{2.17}$$

which relation was to be proved.

2.4 Relations between N and M

It is evident from Eq. (2.6) that if the value of **M** and also its derivative with respect to z is known for a single value of z, then the matrix **N** is also known for that value of z. The converse statement, however, is not at all true. In general, it is necessary to know the value of the **N**-matrix for all values of z lying between z_1 and z_2 if one wishes to determine the matrix **M** which corresponds to that part of the crystal which lies between z_1 and z_2. This is a long-winded way of saying that **M** is an integral function of the **N**-matrix.

Thus, whenever **N** has a prescribed dependence on the coordinate z, it is always possible in principle to express **M** as a function of z. Section 3.0 is devoted to the relations between **N** and **M** when **N** is independent of z, as is the case for a homogeneous crystal.

There is a number of further relations between **N** and **M** which hold quite generally. These relations may be derived from Eq. (2.6) and from the definition of the reciprocal matrix.

Because of the relation

$$\mathbf{M}\mathbf{M}^{-1}=\mathbf{1}, \tag{2.18}$$

Eq. (2.6) may also be written

$$\mathbf{N}=-\mathbf{M}(d\mathbf{M}^{-1}/dz). \tag{2.19}$$

Then by use of Eqs. (2.6) through (2.19), one may also derive

$$\begin{aligned}\mathbf{N}^2&=-(d\mathbf{M}/dz)(d\mathbf{M}^{-1}/dz)\\&=-\mathbf{M}(d\mathbf{M}^{-1}/dz)(d\mathbf{M}/dz)\mathbf{M}^{-1}\\&=\tfrac{1}{2}[\mathbf{M}(d^2\mathbf{M}^{-1}/dz^2)+(d^2\mathbf{M}/dz^2)\mathbf{M}^{-1}].\end{aligned} \tag{2.20}$$

The following relations will also be useful:

$$\begin{aligned}\mathbf{N}^2+(d\mathbf{N}/dz)&=(d^2\mathbf{M}/dz^2)\mathbf{M}^{-1},\\\mathbf{N}^2-(d\mathbf{N}/dz)&=\mathbf{M}(d^2\mathbf{M}^{-1}/dz^2).\end{aligned} \tag{2.21}$$

It is significant to note that all of the relations between **N** and **M** which have been developed so far, as well as others to be developed below, are consistent with the following formal relations

$$\mathbf{M}=\exp(\mathbf{N}z)\quad \mathbf{M}^{-1}=\exp(-\mathbf{N}z), \tag{2.23}$$

provided that the derivatives are defined by

$$d\exp(\mathbf{N}z)/dz\equiv\mathbf{N}\exp(\mathbf{N}z), \tag{2.24}$$

$$d\exp(-\mathbf{N}z)/dz\equiv-\exp(-\mathbf{N}z)\mathbf{N}. \tag{2.25}$$

See also Section 3.2 below.

As a specific example of the meaning of the last three equations, consider the matrix **M** defined by Eq. (2.A). Substitution of Eqs. (2.A) and (2.C) in Eqs. (2.19)–(2.21) then indicates that the equations hold for the matrix defined by Eq. (2.A); in particular one finds:

$$\mathbf{N}^2=-\alpha^2z^2\mathbf{1}, \tag{2.G}$$

$$\begin{aligned}\mathbf{N}^2+\frac{d\mathbf{N}}{dz}&=-\alpha^2z^2\mathbf{1}+\alpha\mathbf{S}(\tfrac{1}{2}\pi)\\&=\begin{pmatrix}-\alpha^2z^2 & -\alpha\\ +\alpha & -\alpha^2z^2\end{pmatrix}.\end{aligned} \tag{2.H}$$

3.0 THE SPECIAL CASE: N INDEPENDENT OF z

This part is devoted to a study of the special case in which **N** is independent of z, as is the case for a homogeneous crystal. The statement that **N** is independent of z means simply that each of the four matrix elements is independent of z.

3.1 Derivation of M in Terms of N by Means of Integration

Perhaps the most straightforward way to determine **M** as a function of z when **N** is independent of z is to integrate Eq. (2.7):

$$d\mathbf{M}/dz=\mathbf{N}\mathbf{M} \tag{3.1}$$

subject to the boundary condition that **M** must be equal to the unit matrix when z is zero. The solution by this means involves the integration of four simultaneous first-order differential equations.

It is possible to obtain the solution more simply, however, by integrating Eq. (2.10), which involves only two simultaneous first-order differential equations. The vector equation (2.10) is equivalent to the following two scalar equations:

$$\begin{aligned}dX/dz&=n_1X+n_4Y,\\dY/dz&=n_3X+n_2Y.\end{aligned} \tag{3.2}$$

These two equations may be combined to obtain the following second-order differential equation which is satisfied by both X and Y:

$$d^2X/dz^2+2T(dX/dz)+DX=0, \tag{3.3}$$

where T is an abbreviation for the half-trace of **N**

$$T\equiv\tfrac{1}{2}(n_1+n_2), \tag{3.4}$$

and where D is the determinant of **N**:

$$D\equiv n_1n_2-n_3n_4. \tag{3.5}$$

It is convenient to introduce at this point also the definition of the discriminant Q:

$$Q\equiv(T^2-D)^{\frac{1}{2}}=(\tfrac{1}{4}(n_1-n_2)^2+n_3n_4)^{\frac{1}{2}}. \tag{3.6}$$

The Eqs. (3.2) and (3.3) must now be integrated subject to the condition that X and Y are equal, respectively, to X_0 and Y_0 when z is equal to zero. The integration may be accomplished by straightforward elementary methods, and the result is the same as that obtained below by a different method and given in Eq. (3.26).

3.2 Series Relation for M in Terms of N

Starting with the relation (3.1), it follows by successive differentiation, in view of the fact that **N** is independent of z, that:

$$d^k\mathbf{M}/dz^k=\mathbf{N}^k\mathbf{M}. \tag{3.7}$$

Substitution of Eq. (3.7) in the Taylor series expansion for **M** in powers of $z-z_0$:

$$\mathbf{M}=\mathbf{M}_0+\sum_{k=1}^{\infty}\frac{d^k\mathbf{M}}{dz^k}\frac{(z-z_0)^k}{k!} \tag{3.8}$$

then leads to:

$$\mathbf{M}=\left\{1+\sum_{k=1}^{\infty}\mathbf{N}^k\frac{(z-z_0)^k}{k!}\right\}\mathbf{M}_0, \tag{3.9}$$

where $\mathbf{M}_0$ is the value of $\mathbf{M}$ at $z=z_0$. This relation may be written symbolically as

$$\mathbf{M}=e^{\mathbf{N}(z-z_0)}\mathbf{M}_0. \tag{3.10}$$

When z_0 is zero, Eq. (3.9) may be written

$$\mathbf{M}=1+\mathbf{N}z+\tfrac{1}{2}\mathbf{N}^2z^2+\tfrac{1}{6}\mathbf{N}^3z^3+\cdots \tag{3.11}$$

which relation may be written symbolically

$$\mathbf{M}=\exp(\mathbf{N}z). \tag{3.12}$$

The last relation should be compared with Eq. (2.23) above.

The conditions for the convergence of these series expansions will not be examined here.

3.3 Identity of the Eigenvectors of N and M

It will now be shown that the eigenvectors of $\mathbf{M}$ are equal to the eigenvectors of $\mathbf{N}$ when $\mathbf{N}$ is independent of z.

The two eigenvectors of $\mathbf{M}$ are defined as the two linearly independent vectors which satisfy the following vector equation:

$$\mathbf{M}\mathcal{E}_M=\lambda_M\mathcal{E}_M, \tag{3.13}$$

where λ_M is a constant for each of the two eigenvectors and is called the eigenvalue corresponding to that eigenvector.

It is physically evident that if $\mathcal{E}_M$ is an eigenvector of a given thickness of a homogeneous crystal, then it is also an eigenvector of any other thickness of that crystal. Accordingly, it may be concluded that $\mathcal{E}_M$ is independent of z when $\mathbf{N}$ is independent of z. By differentiation of Eq. (3.13), it then follows that

$$(d\mathbf{M}/dz)\mathcal{E}_M=(d\lambda_M/dz)\mathcal{E}_M. \tag{3.14}$$

Upon rewriting Eq. (3.13) in the form

$$\mathcal{E}_M=\lambda_M\mathbf{M}^{-1}\mathcal{E}_M, \tag{3.15}$$

and substituting Eq. (3.15) in the left-hand member of Eq. (3.14), one finds:

$$(d\mathbf{M}/dz)\mathbf{M}^{-1}\lambda_M\mathcal{E}_M=(d\lambda_M/dz)\mathcal{E}_M, \tag{3.16}$$

or

$$\mathbf{N}\mathcal{E}_M=(1/\lambda_M)(d\lambda_M/dz)\mathcal{E}_M. \tag{3.17}$$

Since $\mathcal{E}_M$ and $\mathbf{N}$ are both independent of z, it now follows from Eq. (3.17) that $\mathcal{E}_M$ is also an eigenvector of $\mathbf{N}$:

$$\mathcal{E}_M\doteq\mathcal{E}_N, \tag{3.18}$$

where the symbol $\doteq$ means that the quantities on either side are equal except for a constant numerical factor. It then follows also that the coefficient of $\mathcal{E}_M$ on the right-hand side of Eq. (3.17) is the eigenvalue of $\mathbf{N}$ which corresponds to the eigenvector $\mathcal{E}_M$:

$$d\log\lambda_M/dz=\lambda_N. \tag{3.19}$$

Thus, the identity of the eigenvectors has been established.

The relation between the eigenvalues of $\mathbf{M}$ and $\mathbf{N}$ is now easily obtained from Eq. (3.19). By integration one finds

$$\lambda_M=\exp(\lambda_N z), \tag{3.20}$$

where the constant of integration has been evaluated by the condition that λ_M must be unity when $z=0$.

3.4 Explicit Expression for M in Terms of N

An explicit expression for $\mathbf{M}$ in terms of z and the matrix elements of $\mathbf{N}$ will now be obtained on the basis of results derived in the preceding section, when $\mathbf{N}$ is independent of z.

It was shown in Paper IV that if a matrix has the eigenvectors:

$$\mathcal{E}_1=\begin{pmatrix}X_1\\Y_1\end{pmatrix}\quad\mathcal{E}_2=\begin{pmatrix}X_2\\Y_2\end{pmatrix}, \tag{3.21}$$

with the corresponding eigenvalues λ_1 and λ_2, then the matrix can be expressed uniquely in the form:

$$\mathbf{M}=\frac{1}{X_1Y_1-X_2Y_2}\times\begin{pmatrix}\lambda_1X_1Y_2-\lambda_2X_2Y_1 & (\lambda_2-\lambda_1)X_1X_2\\(\lambda_1-\lambda_2)Y_1Y_2 & \lambda_2X_1Y_2-\lambda_1X_2Y_1\end{pmatrix}. \tag{3.22}$$

Also from the material given in Paper IV the eigenvectors of $\mathbf{N}$ and, therefore, also of $\mathbf{M}$, may be expressed in the form

$$\mathcal{E}_N=\mathcal{E}_M=\begin{pmatrix}\frac{1}{2}(n_1-n_2)\pm_1Q_N\\n_3\end{pmatrix}, \tag{3.23}$$

and the eigenvalues of **N** may be expressed in the form

$$\lambda_N = T_N \pm_1 Q_N. \tag{3.24}$$

By substitution of Eq. (3.24) in Eq. (3.20) one finds that the eigenvalues of **N** may be written

$$\lambda_M = \exp((T_N \pm_1 Q_N)z). \tag{3.25}$$

By substitution of Eqs. (3.23) and (3.25) in Eq. (3.22) one then finds the following expression for the matrix **M**:

$$\mathbf{M} = \exp(T_N z)\begin{bmatrix} \cosh Q_N z + \tfrac{1}{2}(n_1 - n_2)\dfrac{\sinh Q_N z}{Q_N} & n_4 \dfrac{\sinh Q_N z}{Q_N} \\ n_3 \dfrac{\sinh Q_N z}{Q_N} & \cosh Q_N z - \tfrac{1}{2}(n_1 - n_2)\dfrac{\sinh Q_N z}{Q_N} \end{bmatrix}. \tag{3.26}$$

This equation is the explicit form of the relation indicated symbolically in Eq. (3.12). It should be noted that the matrix defined by Eq. (3.26) is symmetrical in the parameter Q_N, so that the ambiguity in the sign of Q_N involved in its definition does not introduce any ambiguity in the expression (3.26).

It is of interest to note that this expression has been obtained without any process of integration except that involved in obtaining Eq. (3.20) from Eq. (3.19).

In order to exemplify the results contained in this and the preceding section by a specific case, consider the corresponding **M**- and **N**-matrices defined by Eqs. (2.D) and (2.E):

$$\mathbf{M} = \mathbf{S}(\omega z), \quad \mathbf{N} = \omega \mathbf{S}(\tfrac{1}{2}\pi). \tag{3.A}$$

The eigenvectors of any scalar multiple of a rotation matrix are the vectors which correspond to right and left circularly polarized light. Accordingly, since both **M** and **N** are scalar multiples of rotation matrices, it follows at once that they have the same eigenvectors.

One has from Eqs. (3.4)–(3.6):

$$\begin{aligned} T_M &= \cos\omega, & T_N &= 0, \\ D_M &= 1, & D_N &= \omega, \\ Q_M &= \pm i \sin\omega, & Q_N &= \pm i\omega, \end{aligned} \tag{3.B}$$

from which one finds:

$$\lambda_M = T_M \pm Q_M = \exp(\pm i\omega), \tag{3.C}$$

$$\lambda_N = T_N \pm Q_N = \pm i\omega. \tag{3.D}$$

By comparison of Eq. (3.C) with Eq. (3.D) it is evident that the eigenvectors satisfy the relation (3.20). Finally, by substitution of

$$\begin{aligned} n_1 &= n_2 = 0, \\ n_3 &= -n_4 = \omega \end{aligned} \tag{3.E}$$

in Eq. (3.26), one confirms that the matrix **M** is equal to $\mathbf{S}(\omega z)$.

3.5 Explicit Expression for N in Terms of M

If one knows **M** as a function of z, then the corresponding **N**-matrix can be determined directly from the definition (2.6). If, however, one knows the matrix **M** for a given thickness z but does not know the value of **M** for any other thickness, then it is not possible to determine the matrix **N** uniquely. In the latter case, however, it is possible to provide a multiple-valued solution for **N** in terms of **M**, when **N** is independent of z. The significance of the multiple values will be made clear in a few illustrative examples.

The solution for **N** in terms of **M** will be obtained by solving Eq. (3.26) for the n's in terms of the m's.

First of all, one finds immediately from Eq. (3.26) the following simple expression for the determinant of **M**:

$$D_M = \exp(2T_N z). \tag{3.27}$$

This relation may be inverted to yield T_N as a function of D_M:

$$T_N \equiv \tfrac{1}{2}(n_1 + n_2) = \tfrac{1}{2}\log(\pm_2 D_M^{\frac{1}{2}}), \tag{3.28}$$

where the $\pm_2$ sign has been introduced to give explicit recognition of the fact that the square root of D_M is double valued.

The four scalar equations which are equivalent to the matrix equation (3.26) may be written:

$$T_M \equiv \tfrac{1}{2}(m_1 + m_2) = \exp(T_N z)\cosh Q_N z, \tag{3.29}$$

$$m_1 - m_2 = (n_1 - n_2)\exp(T_N z)\frac{\sinh Q_N z}{Q_N}, \tag{3.30}$$

$$m_3 = n_3 \exp(T_N z)\frac{\sinh Q_N z}{Q_N}, \tag{3.31}$$

$$m_4 = n_4 \exp(T_N z)\frac{\sinh Q_N z}{Q_N}. \tag{3.32}$$

Then from Eqs. (3.29) and (3.27) one finds directly with the help of the definition (3.6):

$$Q_M = \pm_1 \exp(T_N z)\sinh Q_N z, \qquad (3.33)$$

where the $\pm_1$ sign has been introduced because both Q_N and Q_M are ambiguous with regard to sign; once this sign has been specified, however, the same value should be used throughout the calculation. This relation may be rewritten in the form:

$$Q_M = \pm_1 Q_N \exp(T_N z)\frac{\sinh Q_N z}{Q_N}. \qquad (3.34)$$

By comparing Eqs. (3.30), (3.31), (3.32), and (3.34) the following relations may be obtained by inspection:

$$\frac{n_1-n_2}{m_1-m_2} = \frac{n_3}{m_3} = \frac{n_4}{m_4} = \pm_1 \frac{Q_N}{Q_M}. \qquad (3.35)$$

Equations (3.28) and (3.33) are easily combined to yield

$$Q_M = \pm_1 D_M^{\frac{1}{2}} \sinh Q_N z, \qquad (3.36)$$

and this relation becomes upon inversion:

$$\frac{Q_N}{Q_M} = \pm_1 \frac{\sinh^{-1}[Q_M/D_M^{\frac{1}{2}}]}{zQ_M}. \qquad (3.37)$$

The Eqs. (3.35), (3.28), and (3.37) now permit an explicit solution for the n's in terms of the m's. The result may be written in the form:

$$\mathbf{N} = \begin{pmatrix} F_M + \frac{1}{2}(m_1-m_2)G_M & m_4 G_M \\ m_3 G_M & F_M - \frac{1}{2}(m_1-m_2)G_M \end{pmatrix}, \qquad (3.38)$$

where the functions F_M and G_M are defined by:

$$F_M \equiv \frac{1}{z}\log(\pm_2 D_M^{\frac{1}{2}}),$$

$$G_M \equiv \frac{1}{zQ_M}\sinh^{-1}\frac{Q_M}{(\pm_2 D_M^{\frac{1}{2}})} = \frac{1}{zQ_M}\log\frac{Q_M \pm_3 T_M}{(\pm_2 D_M^{\frac{1}{2}})}, \qquad (3.39)$$

where both values of $\pm_2$ and of $\pm_3$ must be used if all of the possible solutions for **M** are to be obtained.

It should be noted that both F_M and G_M are multiple-valued functions.

3.6 Illustrative Examples

In using Eq. (3.38) to determine the matrix **N** when the matrix **M** is known only for a given thickness, the chief difficulty lies in making sure that one has obtained all of the formal solutions given by Eq. (3.38). The best way to make this situation clear is by means of a few examples.

Consider first the matrix

$$\mathbf{M} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}. \qquad (3.40)$$

It is evident upon inspection that this is the matrix of a half-wave plate with its axes parallel with the coordinate axes. The solution of Section 3.5 will now be employed to determine all of the **N**-matrices which are consistent with Eq. (3.40).

One has from Eq. (3.40):

$$\begin{aligned} T_M &= 0, \\ D_M^{\frac{1}{2}} &= 1, \\ Q_M &= i, \end{aligned} \qquad (3.41)$$

where only one of the two possible values of Q_M and of $D_M^{\frac{1}{2}}$ has been specified. Either value of Q_M, if used consistently, leads to the same final result.

In computing the general solution of Eq. (3.38), the only safe procedure is to compute separately the solutions for each of the four different combinations of signs in Eq. (3.39), and then to combine the separate solutions.

In the special case of Eq. (3.40), however, there are only two different choices of sign because T_M is zero. With the first choice, $\pm_2 = +$, one finds

$$\begin{aligned} zF_M &= \log 1 = 2\pi i k, \\ zG_M &= -i\log i = 2\pi l + \tfrac{1}{2}\pi, \end{aligned} \qquad (3.42)$$

whence

$$\mathbf{N}z = \tfrac{1}{2}\pi i \begin{pmatrix} 4k+4l+1 & 0 \\ 0 & 4k-4l-1 \end{pmatrix}, \qquad (3.43)$$

where k and l are any two integers.

With the second choice, $\pm_2 = -$, one finds

$$\begin{aligned} zF_M &= \log(-1) = 2\pi i k + \pi i, \\ zG_M &= -i\log(-i) = 2\pi l - \tfrac{1}{2}\pi, \end{aligned} \qquad (3.44)$$

whence

$$\mathbf{N}z=\tfrac{1}{2}\pi i\begin{pmatrix}4k+4l+1 & 0\\ 0 & 4(k+1)-4l-1\end{pmatrix}. \quad (3.45)$$

Both Eqs. (3.43) and (3.45) are consistent with the following general solution

$$\mathbf{N}z=\begin{pmatrix}\pi i(2p+\frac{1}{2}) & 0\\ 0 & \pi i(2q-\frac{1}{2})\end{pmatrix}, \quad (3.46)$$

where p and q are any two integers.

The case just discussed is one in which the two separate calculations led to the same general solution. Slightly different is the second case:

$$\mathbf{M}=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}. \quad (3.47)$$

The values of T_M, D_M, and Q_M for this matrix are also given by Eq. (3.41); accordingly, the functions F_M and G_M as given by Eq. (3.42) or Eq. (3.44) hold also for this case. One then finds from Eq. (3.38) for the two separate choices of $\pm_2$:

$$\mathbf{N}z=\tfrac{1}{2}\pi\begin{pmatrix}4ki & -4l-1\\ 4l+1 & 4ki\end{pmatrix}, \quad (3.48)$$

$$\mathbf{N}z=\tfrac{1}{2}\pi\begin{pmatrix}(4k+2)i & -4l+1\\ 4l-1 & (4k+2)i\end{pmatrix}, \quad (3.49)$$

where k and l are any two integers.

These two solutions may be combined to obtain the following general solution

$$\mathbf{N}z=\begin{pmatrix}\pi ip & -(2q+p+\frac{1}{2})\pi\\ (2q+p+\frac{1}{2})\pi & \pi ip\end{pmatrix}, \quad (3.50)$$

where p and q are any two integers, and where even values of p correspond to the first choice, and odd values to the second.

The third and last example is one in which there are four different choices of the signs involved in Eq. (3.39). Consider the matrix

$$\mathbf{M}=\tfrac{1}{2}\begin{pmatrix}11 & 5\\ 7 & 9\end{pmatrix}. \quad (3.51)$$

One finds

$$\begin{aligned}Q_M&=3,\\ D_M{}^{\frac{1}{2}}&=4,\\ T_M&=5.\end{aligned} \quad (3.52)$$

Four different solutions will now be found which correspond, respectively, to the following four choices

	$\pm_2$	$\pm_3$
I	+	+
II	+	−
III	−	+
IV	−	−

The functions F_M for choices I and II, and choices III and IV, respectively, are

$$F_M=\log 4=2\epsilon+2\pi ik, \quad (3.53)$$

$$F_M=\log(-4)=2\epsilon+\pi i(2k+1), \quad (3.54)$$

where k is any integer, and ϵ is the principal value of log 2:

$$\epsilon=0.69315\cdots. \quad (3.55)$$

Correspondingly, the functions G_M are given by

$$3zG_M=\log 2=\epsilon+2\pi il, \quad (3.56)$$

$$3zG_M=\log(-\tfrac{1}{2})=-\epsilon+\pi i(2l+1), \quad (3.57)$$

$$3zG_M=\log(-2)=\epsilon+\pi i(2l+1), \quad (3.58)$$

$$3zG_M=\log\tfrac{1}{2}=-\epsilon+2\pi il, \quad (3.59)$$

respectively, where l is any integer.

The four separate solutions for the matrix $\mathbf{N}$ may now be written

$$6\mathbf{N}z=\begin{pmatrix}13\epsilon+2\pi i(6k+l) & 5\epsilon+10\pi il\\ 7\epsilon+14\pi il & 11\epsilon+2\pi i(6k-l)\end{pmatrix}, \quad (3.60)$$

$$6\mathbf{N}z=\begin{pmatrix}11\epsilon+\pi i(12k+2l+1) & -5\epsilon+5\pi i(2l+1)\\ -7\epsilon+7\pi i(2l+1) & 13\epsilon+\pi i(12k-2l-1)\end{pmatrix}, \quad (3.61)$$

$$6\mathbf{N}z=\begin{pmatrix}13\epsilon+\pi i(12k+2l+7) & 5\epsilon+5\pi i(2l+1)\\ 7\epsilon+7\pi i(2l+1) & 11\epsilon+\pi i(12k-2l+5)\end{pmatrix}, \quad (3.62)$$

$$6\mathbf{N}z=\begin{pmatrix}11\epsilon+2\pi i(6k+l+3) & -5\epsilon+10\pi il\\ -7\epsilon+14\pi il & 13\epsilon+2\pi i(6k-l+3)\end{pmatrix}. \quad (3.63)$$

The general solution for **N** is now to be obtained by combining the solutions exhibited in the last four equations.

The desired combination may be expressed as follows:

$$6\mathbf{N}_z = \begin{pmatrix} \epsilon(12+(-1)^p)+\pi i(12r+7q+p) & 5(-1)^p\epsilon+5\pi i(p+q) \\ 7(-1)^p\epsilon+7\pi i(p+q) & \epsilon(12-(-1)^p)+\pi i(12r+5q-p) \end{pmatrix}, \tag{3.64}$$

where p, q, and r are any three integers. The relation between Eq. (3.64) and the partial solutions (3.60)–(3.63) is as follows:

	p	q
I	even	even
II	odd	even
III	even	odd
IV	odd	odd

The physical significance of the solutions obtained in these three examples will be discussed in Section 4.4.

4.0 LAMELLAR REPRESENTATION OF A CRYSTAL

In the latter part (pp. 490–493) of Paper IV a treatment was presented whose object was to show that any crystal without optical activity was optically equivalent to a laminated crystal, whose alternate laminae were identical and identically oriented. One of the alternate laminae was a retardation plate of prescribed retardation and orientation, and the other lamina was a partial polarizer of prescribed orientation and principal absorption coefficients.

One of the purposes of the treatment was to provide a means for the mind to grasp as a whole the optical behavior of such a crystal. To the writer, at least, it was helpful to realize that any section of any homogeneous crystal (with zero optical activity) could be considered, as far as its optical properties for normal incidence were concerned, as an interleaved pile of retardation plates and partial polarizers, in the limit in which the thickness of each lamina approaches zero.

A second purpose of the treatment was to provide a straightforward method of breaking down the complete optical description of the crystal section in terms of the matrix **M**, into a number of relatively simple component properties.

In this section, the treatment will be similar in spirit to that just mentioned, but will be more general in that it will hold for any homogeneous crystal whatever, and different in that the treatment will be based on the **N**-matrices.

Although the treatment given in this section relates explicitly to homogeneous crystals, the method of resolution may also be used for inhomogeneous crystals with the understanding that the resolution holds only at the given point for a given direction of propagation of the light, whereas for homogeneous crystals the resolution holds at every point in the crystal for a given direction of propagation of the light.

4.1 Method of Attack

It is desired to find some method of representing the properties of an arbitrary homogeneous crystal as a combination of a number of simple properties. As a first attempt, one might try to find a simple way of factoring the matrix **M** of the crystal into the product of a finite number of simple **M**-matrices, each of which would represent a simple crystal property, such as circular dichroism, linear birefringence, or isotropic absorption. This effort fails, because the constants which specify the component matrices depend on the order in which the matrices are multiplied.

The resolution of this difficulty may be realized by noting that the **M**-matrix of a very thin section of the crystal differs only slightly from the unit matrix, and that such matrices have a product which is independent of the order in which the matrices are multiplied. Accordingly, it is possible to represent the **M**-matrix of a thin section of the crystal as a product of factors which meaningfully represent the crystal as a combination of simple properties. It turns out that the representation is most conveniently presented in terms of the **N**-matrices.

Consider a sandwich of eight thin laminae. Let the **N**-matrix of each of the laminae be denoted by $\mathbf{N}_k$, where k assumes integral values from one to eight, and let the corresponding

thicknesses of the laminae be τ_k. Each lamina is to be called "thin" if and only if its thickness satisfies the conditions

$$|n_{ik}|\tau_k \ll \tfrac{1}{8}, \tag{4.1}$$

where n_{ik} is the ith matrix element of the kth lamina.

Because of the conditions just stated, Eq. (3.11) may be written

$$\mathbf{M}_k = 1 + \mathbf{N}_k\tau_k + O(\tau_k^2), \tag{4.2}$$

where the term $O(\tau_k^2)$ should be read "a term of the order τ_k^2." Then if the light passes through the laminae in the order in which they are numbered, the **M**-matrix of the sandwich, $\mathbf{M}_s$, is

$$\mathbf{M}_s = \mathbf{M}_8\mathbf{M}_7\cdots\mathbf{M}_2\mathbf{M}_1. \tag{4.3}$$

By use of Eq. (4.2) this may be written:

$$\mathbf{M}_s = 1 + \textstyle\sum_k \mathbf{N}_k\tau_k + O((\sum_k \tau_k)^2). \tag{4.4}$$

It is now convenient to introduce a matrix $\overline{\mathbf{N}}$ which is a weighted average of the $\mathbf{N}_k$'s:

$$\overline{\mathbf{N}} \equiv \frac{\sum_k \mathbf{N}_k\tau_k}{\sum_k \tau_k}. \tag{4.5}$$

Each of the $\mathbf{N}_k$'s is weighted in proportion to the thickness of the lamina. Furthermore, let τ be the total thickness of the sandwich

$$\tau \equiv \textstyle\sum_k \tau_k. \tag{4.6}$$

By the use of the last two abbreviations Eq. (4.4) may be written in the following convenient form:

$$\mathbf{M}_s = 1 + \overline{\mathbf{N}}\tau + O(\tau^2). \tag{4.7}$$

It will now be shown that a pile containing a large number of the octuple sandwiches behaves as though it were a homogeneous crystal with an **N**-matrix equal to $\overline{\mathbf{N}}$, in the limit in which the thickness of the sandwiches approaches zero while the number becomes infinite.

Consider such a pile of the octuple sandwiches. If the number of the sandwiches is q, the total thickness z of the pile is $q\tau$, and the **M**-matrix of the pile is

$$\mathbf{M} = \mathbf{M}_s{}^q = (1 + \overline{\mathbf{N}}\tau + O(\tau^2))^{z/\tau}. \tag{4.8}$$

Now, while the total thickness z of the pile remains constant, let the number q of the octuple sandwiches increase without limit, so that τ approaches zero. Accordingly, one is faced with the problem of calculating the limit of (4.8) as τ approaches zero. This calculation involves finding the qth power of $\mathbf{M}_s$ and may be performed in the manner exemplified at length in Paper III (p. 501); the result of the calculation is that **M** is related to $\overline{\mathbf{N}}$ by Eq. (3.26) provided that $\overline{\mathbf{N}}$ is substituted for **N** in (3.26). Thus in the given limit ($\tau \to 0$) the pile behaves just as though it were a homogeneous crystal with an **N**-matrix equal to $\overline{\mathbf{N}}$.

This result is independent of the ordering of the eight laminae in each sandwich, since no matter what order is used the **M**-matrix of the sandwich is still given by:

$$\mathbf{M}_s = 1 + \overline{\mathbf{N}}\tau + O(\tau^2),$$

as is evident by inspection of the derivation of (4.4) from (4.3).

Thus one has arrived at the concept that any given homogeneous crystal is optically equivalent to a laminated crystal as described above, provided only that the constants of the laminae be so chosen that $\overline{\mathbf{N}}$ is equal to the **N**-matrix of the given crystal. Furthermore, the equivalence is an invariant one in the sense that the equivalence is independent of the ordering of the laminae within each sandwich.

Since eight real constants are required to specify an **N**-matrix (the real and imaginary parts of each of the four matrix elements), it is evident that by suitably choosing the form of each of the eight $\mathbf{N}_k$ matrices so that each contains only one adjustable constant, a unique solution for the constants may always be obtained.

In the next section the form of the eight $\mathbf{N}_k$ matrices will be chosen as just described, and furthermore in such a way that each corresponds to a particularly simple type of optical behavior. For example, the matrix $\mathbf{N}_3$ will represent circular birefringence; $\mathbf{N}_5$ will represent linear birefringence with axes parallel to the coordinate axes, etc.

Before specifying the $\mathbf{N}_k$'s, however, it is convenient to introduce a simplification and a change in notation. Henceforth, it will be assumed that all of the laminae have the same thickness:

$$\tau_k = \tfrac{1}{8}\tau.$$

Furthermore, let $\mathbf{\Theta}_k$ be defined as one-eighth of $\mathbf{N}_k$:

$$\mathbf{\Theta}_k \equiv \tfrac{1}{8}\mathbf{N}_k.$$

One then has

$$\bar{\mathbf{N}} = \sum_k \mathbf{\Theta}_k.$$

The $\mathbf{\Theta}_k$'s are such that if every $\mathbf{\Theta}_k$ is zero except one, denoted by $\mathbf{\Theta}_i$, then in the limit ($\tau \to 0$) the pile of octuple sandwiches will behave as though it were a homogeneous crystal with a N-matrix equal to $\mathbf{\Theta}_i$. Thus it seems reasonable to postulate that the magnitude of each $\mathbf{\Theta}_k$ represents the amount of the corresponding optical property in a homogeneous crystal whose N-matrix is $\sum_k \mathbf{\Theta}_k$. The examination of several M-matrices from this point of view is carried out in Section 4.4.

4.2 Specification of the Eight Θ Matrices

The eight matrices denoted by $\mathbf{\Theta}_k$ in the preceding section will now be determined so that each of them represents a particularly simple type of crystal. The eight matrices are listed in Table I along with a description of the corresponding optical behavior and the quantitative significance of the associated constants. For example, $\mathbf{\Theta}_1$ introduces a progressive change in phase corresponding to the mean index of refraction of the material, $\mathbf{\Theta}_2$ introduces an absorption corresponding to the mean absorption coefficient of the material, $\mathbf{\Theta}_3$ introduces circular birefringence (optical activity), and so on. It will be noted that each of the matrices involves a single real number so that, as must be the case, the matrix $\bar{\mathbf{N}}$ involves eight independent real numbers.

A number of conventions are involved in Table I. The time factor, the index of refraction n, and the extinction coefficient k are as they appear in the following relation for an isotropic medium:

$$\mathcal{E} = \mathcal{E}_0 \exp(i\omega t - 2\pi(k+in)z/\lambda).$$

The x, y, and z coordinate system is assumed to be right handed, and the light to travel in the direction of the positive z axis. Positive rotations about the z axis are defined as those which rotate from the positive x axis toward the positive y axis. Then the vectors representing right and left circularly polarized lights are*

$$\mathcal{E}_r = \begin{pmatrix} -i \\ 1 \end{pmatrix},$$

$$\mathcal{E}_l = \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

In a stationary plane normal to the z axis, circularly polarized light whose vector rotates in the positive direction is left circularly polarized. Furthermore, with these conventions a rotation of the plane of polarization in the positive direction is a levorotation.

A rather unusual type of resolution is involved in the last four of the matrices listed in Table I: The linear birefringence is split into two components, one that part which is parallel with the coordinate axes, and the other parallel with the bisectors of the coordinate axes. The sum of $\mathbf{\Theta}_5$ and $\mathbf{\Theta}_7$ is

$$\mathbf{\Theta}_5 + \mathbf{\Theta}_7 = \begin{pmatrix} ig_0 & ig_{45} \\ ig_{45} & -ig_0 \end{pmatrix}. \tag{4.9}$$

If one compares the last expression with the matrix of a linearly birefringent plate with its fast axis rotated through the angle α from the x axis:

$$\begin{pmatrix} ig\cos 2\alpha & ig\sin 2\alpha \\ ig\sin 2\alpha & -ig\cos 2\alpha \end{pmatrix}, \tag{4.10}$$

where g is positive, it follows that (4.9) is equivalent to a wave plate of retardation specified by

$$g^2 = g_0^2 + g_{45}^2, \tag{4.11}$$

and with a fast axis at the angle

$$\tan 2\alpha = g_{45}/g_0. \tag{4.12}$$

* In the writer's opinion *the* definition of right circularly polarized light is that an instantaneous picture of the space distribution of its electric vector represents a right spiral. This definition is adhered to by Pockels (*Lehrbuch der Kristalloptik* (B. G. Teubner, Berlin, 1906), pp. 7–8) and by Born (*Optik* (Verlag Julius Springer, Berlin, 1933), p. 24), and is the convention in terms of which all of the experimental data on optical activity is presented in the literature. Born, however, seems dissatisfied with the convention, as indicated by the following quotation: "Gemäss der Tradition, an die man sich halten muss, bezieht man den Drehsinn 'rechts' und 'links' nicht auf die Fortpflanzungsrichtung, sondern auf die entgegengesetzte, die 'Blickrichtung'." In the opinion of the writer, however, the definition should not be founded upon the time variation of the electric vector in a given plane as seen from a given direction, but should rather be founded on the spatial configuration of the electric vector at a given moment.

TABLE I.

This table shows the form of the $\mathbf{N}$-matrices for eight different and independent types of crystalline behavior.

$\Theta_1=-\eta\begin{pmatrix}i & 0\\ 0 & i\end{pmatrix}$	The parameter η is the propagation constant, or the phase retardation per unit thickness, and is thus related to the index of refraction n by $\eta=2\pi n/\lambda$.
$\Theta_2=-\kappa\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$	The parameter κ is the amplitude absorption coefficient to the base e per unit thickness, and is thus related to the extinction coefficient k by $\kappa=2\pi k/\lambda$.
$\Theta_3=\omega\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$	The parameter ω is a measure of the circular birefringence, and is equal to the rotation (in the positive direction) of the plane of linearly polarized light, in radians per unit thickness. It is equal to half of the difference of propagation constants for right and left circularly polarized lights, $\omega=\frac{1}{2}(\eta_r-\eta_l)$, and is positive for crystals which are levorotatory.
$\Theta_4=\delta\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}$	The parameter δ is a measure of the circular dichroism, and is equal to half of the difference of the absorption coefficients for left and right circularly polarized lights: $\delta=\frac{1}{2}(\kappa_l-\kappa_r)$. The parameter is positive for crystals which are more transparent for right polarized light.
$\Theta_5=g_0\begin{pmatrix}i & 0\\ 0 & -i\end{pmatrix}$	The parameter g_0 is a measure of that part of the linear birefringence which is parallel with the coordinate axes. It is equal to one-half of the difference between the two principal propagation constants $g_0=\frac{1}{2}(\eta_y-\eta_x)$, and is thus positive when the fast (smaller index) axis is parallel with the x axis.
$\Theta_6=p_0\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$	The parameter p_0 is a measure of that part of the linear dichroism which is parallel with the coordinate axes. It is equal to one-half of the difference of the two principal absorption coefficients $p_0=\frac{1}{2}(\kappa_y-\kappa_x)$, and is thus positive when the more highly transmitting axis is parallel with the x axis.
$\Theta_7=g_{45}\begin{pmatrix}0 & i\\ i & 0\end{pmatrix}$	The parameter g_{45} is a measure of that part of the linear birefringence which is parallel with the bisectors of the coordinate axes. It is equal to one-half of the difference between the two principal propagation constants $g_{45}=\frac{1}{2}(\eta_{-45}-\eta_{45})$, and is thus positive when the fast axis bisects the positive x and y axes.
$\Theta_8=p_{45}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$	The parameter p_{45} is a measure of that part of the linear dichroism which is parallel with the bisectors of coordinate axes. It is equal to one-half of the difference between the two principal absorption coefficients $p_{45}=\frac{1}{2}(\kappa_{-45}-\kappa_{45})$, and is thus positive when the more highly transmitting axis bisects the positive x and y axes.

The linear dichroism is correspondingly expressed by Θ_6 and Θ_8.

$$\overline{\mathbf{N}}=\begin{pmatrix}-\kappa+p_0-i\eta+ig_0 & -\omega+p_{45}-i\delta+ig_{45}\\ \omega+p_{45}+i\delta+ig_{45} & -\kappa-p_0-i\eta-ig_0\end{pmatrix}. \tag{4.13}$$

It is evident from the form of this matrix that $\overline{\mathbf{N}}$ may be made equal to any given $\mathbf{N}$-matrix by suitable choice of the eight constants, and that this choice is always unique.

From this expression for $\overline{\mathbf{N}}$ one finds at once:

$$T_{\overline{N}}=-\kappa-i\eta,$$
$$Q_{\overline{N}}^2=(p_0+ig_0)^2+(p_{45}+ig_{45})^2-(\omega+i\delta)^2. \tag{4.14}$$

By Table I and Eq. (4.7), the matrix $\overline{\mathbf{N}}$ may be written explicitly as follows:

Thus, $T_{\overline{N}}$ and $Q_{\overline{N}}$ together depend on all of the eight parameters.

The general expression for the matrix $\overline{\mathbf{M}}$ which corresponds to $\overline{\mathbf{N}}$ may be obtained by substituting Eqs. (4.13) and (4.14) in Eq. (3.26).

4.3 Special Case of Non-Absorbing Crystals

In the special case in which the crystal does not absorb, the even numbered Θ_k's will all be

zero, so that the matrix $\overline{\mathbf{N}}$ may be written in the following relatively simple form:

$$\overline{\mathbf{N}}=\begin{pmatrix}-i\eta+ig_0 & -\omega+ig_{45}\\ \omega+ig_{45} & -i\eta-ig_0\end{pmatrix}, \tag{4.15}$$

from which one finds

$$\begin{aligned}T_{\overline{N}}&=-i\eta,\\ Q_{\overline{N}}{}^2&=-(g^2+\omega^2)=-\Gamma^2,\end{aligned} \tag{4.16}$$

where g is defined by Eq. (4.11). Substitution of Eqs. (4.15) and (4.16) in Eq. (3.26) then yields:

$$\overline{\mathbf{M}}=\exp(-i\eta z)\begin{bmatrix}\cos\Gamma z+ig_0\dfrac{\sin\Gamma z}{\Gamma} & (-\omega+ig_{45})\dfrac{\sin\Gamma z}{\Gamma}\\ (\omega+ig_{45})\dfrac{\sin\Gamma z}{\Gamma} & \cos\Gamma z-ig_0\dfrac{\sin\Gamma z}{\Gamma}\end{bmatrix}. \tag{4.17}$$

Perhaps the most important simple observation that can be made about this unitary matrix is that, except for the factor written before the matrix which represents the absolute phase, the matrix $\mathbf{M}$ is periodic in z with the period:

$$z_0=2\pi/(g^2+\omega^2)^{\frac{1}{2}}=2\pi/\Gamma. \tag{4.18}$$

This fact can also be represented by stating that the wave number (the number of periods in unit length) is given by:

$$k=1/z_0=(g^2+\omega^2)^{\frac{1}{2}}/2\pi. \tag{4.19}$$

If one introduces the wave numbers corresponding to g and ω when either of these types of retardation exists alone,

$$k_g=g/2\pi;\quad k\omega=\omega/2\pi, \tag{4.20}$$

one has

$$k=(k_g{}^2+k_\omega{}^2)^{\frac{1}{2}}. \tag{4.21}$$

This relation indicates that a small amount of circular birefringence superposed on a large amount of linear birefringence will affect the value of the wave number k only by a quantity of the second order in their ratio.

The existence of the periodic behavior exhibited by Eq. (4.17) can also be predicted by the representation involving Poincaré's sphere.[2] In this representation the distance z_0 is the distance in which the sphere experiences a rotation through 360°.

[2] H. Poincaré, *Théorie Mathématique de la Lumière* (Gauthier–Villars et Fils, Paris, 1892).

4.4 Illustrative Examples

The formulae developed in this section will now be used to discuss the physical significance of the mathematical solutions obtained in Section 3.6.

Consider the first example, in which it is found that Eq. (3.46) represented the most general value of $\mathbf{N}$ which is consistent with the $\mathbf{M}$-matrix defined by Eq. (3.40). By comparing Eq. (3.46) with the general expression (4.13), one finds

$$\begin{aligned}\eta&=-\pi(p+q)/z,\\ g_0&=\pi(p-q+\tfrac{1}{2})/z,\end{aligned} \tag{4.22}$$

with the other six constants all zero. Thus the crystal is linearly birefringent with its axes parallel to the coordinate axes. It is an $(n+\frac{1}{2})$-wave plate, where n is even if $p+q$ is even, and odd otherwise. The quantity η is the propagation constant the crystal would have if it were replaced with an isotropic material with an index of refraction equal to the mean of the two principal indices of the crystal. The number of wave-lengths within the thickness of the isotropic material would be equal to $-\frac{1}{2}(p+q)$.

The second example is represented by Eqs. (3.47) and (3.50). One finds

$$\begin{aligned}\eta&=-\pi p/z,\\ \omega&=(2q+p+\tfrac{1}{2})\pi/z,\end{aligned} \tag{4.23}$$

with the other six constants all zero. Thus the crystal is optically active, and rotates the plane 90°, plus or minus a number of complete revolutions if p is even, and plus or minus an odd number of 180° rotations if p is odd.

The last example is much more complex. By comparing the general solution (3.64) with Eq.

(4.13), one finds

$$\begin{aligned} \eta &= -\pi(q+2r)/z, \\ \kappa &= -2\epsilon/z, \\ \omega &= \tfrac{1}{8}(-1)^p\epsilon/z = p_0, \\ \delta &= \tfrac{1}{8}\pi(p+q)/z = g_0, \\ g_{45} &= \pi(p+q)/z, \\ p_{45} &= (-1)^p\epsilon. \end{aligned} \tag{4.24}$$

Thus, in general, none of the constants is zero. In the simplest possible case, in which p, q, and r are all zero, the constants become

$$\begin{aligned} -\tfrac{1}{2}\kappa = 6\omega = 6p_0 = p_{45} = \epsilon, \\ \eta = \delta = g_0 = g_{45} = 0. \end{aligned} \tag{4.25}$$

In this simple case, the crystal is described as a combination of a negative isotropic absorption, a positive optical activity, and a linear dichroism with its fast axis at the angle defined by

$$\tan 2\alpha = p_{45}/p_0 = 6, \tag{4.26}$$

or

$$\alpha = 40.8°. \tag{4.27}$$

The circular dichroism, and the linear birefringence are both zero.

The only one of the eight constants which is independent of p, q, and r is κ, and its value is $-1.385/z$. Since a physically possible crystal cannot have a negative absorption coefficient, the matrix defined by Eq. (3.51) does not represent a physically realizable crystal. The last statement is easily confirmed by substitution of Eq. (3.51) in Eq. (31) of Paper IV, which substitution leads to a value of Γ (as defined in Paper IV) of 8.4. It follows that the matrix (3.51) must be divided by a number whose absolute value is not less than 8.4 in order to convert it to a matrix which represents a physically realizable crystal.

5.0 THE TWISTED CRYSTAL

5.1 The Uniformly Twisted Crystal

Sections 3.0 and 4.0 have treated the important special case in which $\mathbf{N}$ is independent of z. There is another special dependence of $\mathbf{N}$ upon z which permits a simple solution. This second special case is the case in which an originally homogeneous crystal is twisted uniformly about an axis parallel to the direction of transmission of the light.

Let $\mathbf{N}_0$ be the matrix corresponding to the untwisted crystal. Then, according to Eq. (2.17), the matrix $\mathbf{N}$ which corresponds to the twisted crystal is given by

$$\mathbf{N} = \mathbf{S}(kz)\mathbf{N}_0\mathbf{S}(-kz), \tag{5.1}$$

where $\mathbf{S}$ is the rotation matrix defined by Eq. (2.15) and where k is the angular twist per unit thickness. The differential equation (2.10) may now be written

$$\frac{d\mathcal{E}}{dz} = \mathbf{S}(kz)\mathbf{N}_0\mathbf{S}(-kz)\mathcal{E}. \tag{5.2}$$

The general solution of this differential equation will now be found.

Let $\mathcal{E}'$ be defined by

$$\mathcal{E}' \equiv \mathbf{S}(-kz)\mathcal{E}. \tag{5.3}$$

Substitution of this relation in Eq. (5.2) then yields

$$\mathbf{S}(kz)\frac{d\mathcal{E}'}{dz} + \frac{d\mathbf{S}(kz)}{dz}\mathcal{E}' = \mathbf{S}(kz)\mathbf{N}_0\mathcal{E}', \tag{5.4}$$

or

$$d\mathcal{E}'/dz = \{\mathbf{N}_0 - k\mathbf{S}(\tfrac{1}{2}\pi)\}\mathcal{E}', \tag{5.5}$$

where the relation

$$(d/dz)\mathbf{S}(kz) \equiv \omega\mathbf{S}(kz)\mathbf{S}(\tfrac{1}{2}\pi) \tag{5.6}$$

has been used. The substitution

$$\mathbf{N}' \equiv \mathbf{N}_0 - k\mathbf{S}(\tfrac{1}{2}\pi) \tag{5.7}$$

then permits Eq. (5.5) to be written in the folfowing form:

$$d\mathcal{E}'/dz = \mathbf{N}'\mathcal{E}', \tag{5.8}$$

where $\mathbf{N}'$ is a matrix which is independent of z. Accordingly, the solution of Eq. (5.8) may be written down directly from the solution obtained in Section 3.4:

Let the relation (3.26) by which $\mathbf{M}$ is expressed in terms of $\mathbf{N}$ by written symbolically as

$$\mathbf{M} = \exp(\mathbf{N}z) \tag{5.9}$$

in accordance with Eq. (3.12). Then the solution of (5.8) may be written

$$\mathcal{E}' = \exp(\mathbf{N}'z)\mathcal{E}_0', \tag{5.10}$$

where $\mathcal{E}_0'$ is the value of the vector $\mathcal{E}_0'$ at $z=0$, as in Section 2.1. By the use of (5.3), the last equation becomes

$$\mathbf{S}(-kz)\mathcal{E} = \exp(\mathbf{N}'z)\mathcal{E}_0, \tag{5.11}$$

or

$$\mathcal{E} = \mathbf{S}(kz)\exp(\mathbf{N}'z)\mathcal{E}_0, \tag{5.12}$$

from which one infers that the matrix $\mathbf{M}$ which

corresponds to the **N**-matrix (5.1) is

$$\mathbf{M}=\mathbf{S}(kz)\exp(\{\mathbf{N}_0-k\mathbf{S}(\tfrac{1}{2}\pi)\}z). \quad (5.13)$$

Thus the matrix **M** of the twisted crystal is equal to $\mathbf{S}(kz)$ multiplied by the **M**-matrix corresponding to a homogeneous crystal with an N-matrix equal to $\mathbf{N}_0-k\mathbf{S}(\frac{1}{2}\pi)$.

Equation (5.13) is the general integral of Eq. (5.2) which satisfies the boundary condition that **M** is equal to the unit matrix when z is equal to zero.

5.2 The Arbitrarily Twisted Crystal

The transformation effected in the preceding section may also be employed with an arbitrarily twisted crystal. If $\mathbf{N}_0$ is the matrix of the untwisted crystal, the **N** matrix of the twisted crystal is

$$\mathbf{N}=\mathbf{S}(\omega(z))\mathbf{N}_0\mathbf{S}(-\omega(z)), \quad (5.14)$$

where $\omega(z)$ is the arbitrary function of z which specifies the angle of twist at the coordinate z.

Then by the same transformation used in Section 5.1, one finds that the **M** matrix which corresponds to Eq. (5.14) is given by

$$\mathbf{M}=\mathbf{S}(\omega(z))\mathbf{M}', \quad (5.15)$$

where $\mathbf{M}'$ is the **M** matrix which corresponds to

$$\mathbf{N}'\equiv\mathbf{N}_0-(d\omega(z)/dz)\mathbf{S}(\tfrac{1}{2}\pi). \quad (5.16)$$

Only in the case of the uniformly twisted crystal is $d\omega(z)/dz$ a constant, and thus only in this case is the matrix $\mathbf{N}'$ independent of z.

In general, however, it is easier to determine the matrix $\mathbf{M}'$ than it is to determine the matrix **M** directly from Eq. (5.14), since the trigonometrical factors which occur in **N** do not occur in $\mathbf{N}'$.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 8 AUGUST, 1948

The Multipupil in Phase Microscopy*

HAROLD OSTERBERG
American Optical Company, Scientific Instrument Division, Buffalo 15, New York

(Received March 3, 1948)

Auxiliary lenses may be employed in a modified phase microscope to form additional images of the condenser diaphragm. The images of the condenser diaphragm are called multipupils. They can serve to increase the flexibility of the phase microscope because a diffraction plate may be placed at any one or at all of the multipupils. Multipupil systems serve as effective means for phase microscopy. The bipupil system is of most practical interest because it combines the required flexibility with simplicity. It is shown on general theoretical grounds that the multipupil system is equivalent to a single system of phase microscopy whose over-all magnification is equal to the product of the magnifications of the multipupil system and whose pupil function is the product of the complex pupil functions associated with each of the multipupils. The physical interpretation of the composite pupil function as a product is that the optical paths through corresponding areas of the multipupils are additive, while the corresponding amplitude transmissions are multiplied. The method of proof serves to extend the theory of phase microscopy to multipupil systems.

I. INTRODUCTION

THE diffraction plate of the conventional phase microscope is located, awkwardly for some purposes, at or within the objective. The diffraction plate acts as an exit pupil. It is conjugate to a specialized condenser diaphragm. Some applications of the phase microscope require at least one intermediate, accessible, real image of the exit pupil. The resulting series of exit pupils is called a multipupil. Multipupil systems function as effective means for phase microscopy. They possess additional degrees of freedom in that diffraction plates may be located at any one or at all of the multipupils.

II. BIPUPIL SYSTEMS

The essential physical principles of multipupil systems are involved in the simpler but highly flexible bipupil system. It will be shown that the theory of phase microscopy can be extended

* Presented at the meeting of the Optical Society of America in New York, New York, 1948.

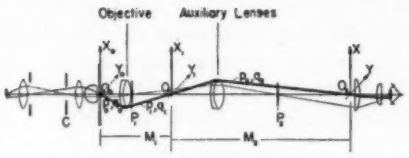


FIG. 1. A bipupil system for phase microscopy.

with slight modification to include bi or multi-pupil systems.

Figure 1 indicates how the condenser diaphragm, C, is focused successively at the bipupil positions P_1 and P_2. A diffraction plate may be placed at either P_1 or P_2 or at both P_1 and P_2. P_2 is the more accessible position. The object plane O_0 is focused at the plane O_1 and again at the plane O_2 of the eyepiece. This constitutes a simplified system which is, however, complex enough to include all of the essential characteristics of bipupil systems.

The coordinates of the planes O_0, O_1, and O_2 are designated as X_0, Y_0, X_1, Y_1 and X, Y, respectively. M_1 and M_2 are magnifications between the indicated pairs of these planes. The heavily drawn ray is an axial ray traced from the origin O_0 of the object plane. Its optical direction cosines are denoted, as indicated, by p_0, q_0, p_1, q_1 and p_2, q_2. The complex amplitude and phase transmissions of the diffraction plates at P_1 and P_2 can be specified by means of these optical direction cosines as $P_1(p_1, q_1)$ and $P_2(p_2, q_2)$, respectively. Similarly, the amplitude and phase transmission of the object plane is specified by $f(x_0, y_0, p_0, q_0)$. This function is called the object function. It is considered to be zero beyond the field of view of the optical system. If the objective or auxiliary lenses have spherical aberration or if they subject the axial bundle of rays to variations due to Fresnel reflection and refraction phenomena, the corresponding effects will be considered to be included in the pupil functions $P_1(p_1, q_1)$ and $P_2(p_2, q_2)$. It is assumed that Abbe's sine condition is satisfied throughout the image forming system.

Let the complex function $H(x_1, y_1, p_0, q_0)$ denote the amplitude and phase distribution produced over the intermediate image plane X_1, Y_1 by a single incident wave front whose normal has the optical direction cosines p_0, q_0. It can be shown[1] that

$$H(x_1, y_1, p_0, q_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_0, y_0, p_0, q_0) \times \exp(2\pi i(p_0x_0+q_0y_0)) \times U_1(x_1-M_1x_0, y_1-M_1y_0)dx_0dy_0. \quad (1)$$

Since the object function is considered to be zero beyond the field of view of the optical system, it is permissible to assign the indicated, infinite limits of integration. All distances are measured in terms of the wave-length of the incident, monochromatic beam of light. The complex function $U_1(x_1, y_1)$ and the pupil function $P_1(p_1, q_1)$ are related by the pair of Fourier transforms:

$$U_1(x_1, y_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_1(p_1, q_1) \times \exp(2\pi i(p_1x_1+q_1y_1))dp_1dq_1; \quad (2)$$

$$P_1(p_1, q_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_1(x_1, y_1) \times \exp(-2\pi i(p_1x_1+q_1y_1))dx_1dy_1. \quad (3)$$

Infinite limits of integration are assigned to Eq. (2) in view of the consideration that

$$P_1(p_1, q_1) = 0 \quad \text{when} \quad p_1^2+q_1^2 > n_1^2 \sin^2\vartheta_1; \quad n_1 = 1; \quad (4)$$

where $\sin\vartheta_1$ is the numerical aperture of the objective, Fig. 1, with respect to its image space.

The function $H(x_1, y_1, p_0, q_0)$, corresponding to the intermediate image plane X_1, Y_1, becomes the object function for the system of auxiliary lenses, Fig. 1. It can be shown[2] in a manner

[1] Bennett, Jupnik, Osterberg, and Richards, Trans. Am. Micro. Soc. 65, 107–108 (1946).

[2] Proof: Note that since $U_2(x, y)$ is given by Eq. (6) the weight factor $U_2(x-M_2x_1, y-M_2y_1)$ of Eq. (5) has the physical interpretation of the diffraction pattern produced by the auxiliary lens system about the point $x=M_2x_1$, $y=M_2y_1$ in the image plane X, Y from a point particle whose amplitude and phase is taken as unity and which is located at x_1, y_1 in the X_1, Y_1 plane. Because this point particle has the phase and amplitude given by the vector or complex number $H(x_1, y_1, p_0, q_0)$, the product $H(x_1, y_1, p_0, q_0)U_2(x-M_2x_1, y-M_2y_1)$ determines the actual phase and amplitude distribution around the point $x=M_2x_1$, $y=M_2y_1$ from the single point particle located at x_1, y_1. The amplitude and phase distribution $H(x_1, y_1, p_0, q_0)$ is coherent because it arises from a single, incident wave

similar to the derivation of Eq. (1) that the distribution $F_0(x, y, p_0, q_0)$ of amplitude and phase over the image plane X, Y is given by

$$F_0(x, y, p_0, q_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x_1, y_1, p_0, q_0) \times U_2(x - M_2x_1, y - M_2y_1)dx_1dy_1. \quad (5)$$

This distribution is due to a single incident wave front whose normal has the optical direction cosines p_0, q_0. Optical systems do not conform accurately to the requirements of Eq. (5) in the matter of the assigned infinite limits. Because all distances are to be measured in wave-lengths, the field of view of the usual optical system extends over a great number of wave-lengths. For such systems $F_0(x, y, p_0, q_0)$, as computed from Eq. (5), will agree very closely with the actual distribution of amplitude and phase over the X, Y plane. It is assumed, accordingly, that the field of view of the optical system extends over a great number of wave-lengths. The complex function $U_2(x, y)$ and the pupil function $P_2(p_2, q_2)$ are related as the pair of Fourier transforms:

$$U_2(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_2(p_2, q_2) \times \exp(2\pi i(p_2x + q_2y))dp_2dq_2; \quad (6)$$

$$P_2(p_2, q_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U_2(x, y) \times \exp(-2\pi i(p_2x + q_2y))dxdy; \quad (7)$$

$$P_2(p_2, q_2) = 0 \quad \text{when} \quad p_2^2 + q_2^2 > n_2^2 \sin^2\vartheta_2; \quad n_2 = 1; \quad (8)$$

in which $\sin\vartheta_2$ is the numerical aperture of the auxiliary lens system, Fig. 1, with respect to its image space.

By introducing $H(x_1, y_1, p_0, q_0)$ from Eq. (1) into Eq. (5) and by making the changes of variable

$$x - M_2x_1 \equiv x'; \quad y - M_2y_1 \equiv y'; \quad (9)$$

and

$$M \equiv M_1M_2; \quad (10)$$

one finds in a straightforward manner that

$$F_0(x, y, p_0, q_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_0, y_0, p_0, q_0) \times \exp(2\pi i(p_0x_0 + q_0y_0)) \times U(x - Mx_0, y - My_0)dx_0dy_0, \quad (11)$$

in which

$$U(x - Mx_0, y - My_0) \equiv \frac{1}{M_2^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \times U_1\left(\frac{x - Mx_0 - x'}{M_2}, \frac{y - My_0 - y'}{M_2}\right) \times U_2(x', y')dx'dy'. \quad (12)$$

Since U_1 and U_2 are double integrals, $F_0(x, y, p_0, q_0)$, Eq. (11), is an octuple integral. It is reducible, however, to a quadruple integral, Eq. (14), in the following manner.

The right-hand member of Eq. (12) consists of a combination of convolution[3,4] of the functions U_1 and U_2 together with a change of scale[5] in the function U_1. In view of the pairs of Fourier transformations of Eqs. (2), (3), (6), and (7) and the laws which relate convolution and change of scale in direct and inverse Fourier transforms, it follows that

$$U(x - Mx_0, y - My_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P_1(M_2p, M_2q) \times P_2(p, q) \exp(2\pi i[p(x - Mx_0) + q(y - My_0)])dpdq. \quad (13)$$

front whose normal has the optical direction cosines p_0, q_0. Considering a point x, y in the image plane, the resultant amplitude and phase at this point is, therefore, given by the vector sum of the overlapping vector quantities $H(x_1, y_1, p_0, q_0) \times U_2(x - M_2x_1, y - M_2y_1)$ corresponding to all points x_1, y_1 in the intermediate image plane X_1, Y_1. Thus sum is readily seen to be the integral of Eq. (5).

The same reasoning may be applied to deduce Eq. (1). In this case the object particles correspond to the points x_0, y_0 whose amplitude and phase are given by the product $f(x_0, y_0, p_0, q_0) \cdot \exp(2\pi i(p_0x_0 + q_0y_0)$ in which the exponential phase factor determines the phase variation produced over the object plane by the inclined, incident wave front whose normal has the optical direction cosines p_0, q_0. The elementary diffraction function $U_1(x_1 - M_1x_0, y_1 - M_1y_0)$ refers this time to the objective, Fig. 1, and the conjugate planes X_0, Y_0 and X_1, Y_1.

[3] Churchill, *Modern Operational Mathematics in Engineering*, (McGraw-Hill Book Company, Inc., 1944), pp. 36–37. The laws governing convolution and change of scale for Fourier transforms are the same as those for Laplace transforms. Adequate references to the well-known double integral are so unavailable that the reader must make his own extension from the single to the double integral.

[4] Campbell and Foster, *Fourier Integrals and Practical Applications*, Bell Telephone System Technical Publications, Monograph B-584 (1931), pp. 37 and 39.

[5] See reference 3, p. 23, Eq. (3).

In summary, for a bipupil system:

$$F_0(x, y, p_0, q_0) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_0, y_0, p_0, q_0) \times \exp(2\pi i(p_0x_0+q_0y_0)) \times U(x-Mx_0, y-My_0)dx_0dy_0; \quad (14)$$

$$U(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(p, q) \times \exp(2\pi i(px+qy))dpdq; \quad (15)$$

$$P(p, q) \equiv P_1(M_2p, M_2q)P_2(p, q); \quad (16)$$

$$M = M_1M_2; \quad (17)$$

$$P(p, q) = 0 \text{ when } p^2+q^2 > n^2 \sin^2\vartheta_m; \; n = 1. \quad (18)$$

The direction cosines p, q belong to the converging axial bundle in the image space of the auxiliary lens system, Fig. 1. The numerical aperture of this axial bundle of rays is redefined as $\sin\vartheta_m$. The functions $U(x, y)$ and $P(p, q)$ form a pair of Fourier transforms which relate to the combined optical system.

III. PHYSICAL INTERPRETATION

Equations (14), (15), and (18) are identical in form with those which govern the phenomena of phase microscopy in the conventional phase microscope. The bipupil system can, therefore, be regarded as a single system whose pupil function $P(p, q)$ is given by Eq. (16) and whose over-all magnification $M = M_1M_2$. With reference to Eq. (16) the coordinates M_2p, M_2q and p, q in P_1 and P_2, respectively, correspond to the points of intersection of the incident, axial ray p_0, q_0 with the plane of the diffraction plates at the bipupils. The interpretation of the product $P_1(M_2p, M_2q)P_2(p, q)$ of the complex pupil functions is, therefore, that the optical paths through corresponding areas of the bipupils are additive, whereas the amplitude transmissions are multiplied.

In a well corrected lens system there is a one to one correspondence between the incident rays with optical direction cosines p_0, q_0 and the points in the source of light. Consequently, Eqs. (14) to (18) apply to each point in an extended source of light. The energy density produced on the X, Y plane, Fig. 1, by each point in the source of light is, therefore, given by $|F_0(x, y, p_0, q_0)|^2$. Upon the usual assumption that each point in the source of light behaves as an independent oscillator, the total energy distribution in the image plane X, Y is determined by the law of superposition of $|F_0(x, y, p_0, q_0)|^2$ for all contributing points in the source of light, Consequently, the interpretation of the above paragraph applies also to the total energy density $G(x, y)$ in the image plane. Explicitly, $G(x, y)$ is determined by the integral

$$G(x, y) = \iint \frac{S(p_0, q_0)}{n_0^2-p_0^2-q_0^2} \times |F_0(x, y, p_0, q_0)|^2 dp_0dq_0, \quad (19)$$

in which $S(p_0, q_0)$ measures the strength of the source of light, n_0 is the refractive index in the object space, and the domain of integration extends over the contributing portion of the source of light, that is, over the rays p_0, q_0 which penetrate the condenser diaphragm.

IV. EXTENSION TO MULTIPUPIL SYSTEMS

All of the conclusions which have been stated for the bipupil system apply with slight modification to multipupil systems. When more than two pupils are considered, a longer sequence of Fourier transforms arises. If the Abbe sine condition is obeyed in the formation of all of the intermediate images of the object, it is always possible to obtain the solution in the form of Eqs. (14), (15), and (18). Corresponding to N multipupils

$$P(p, q) = P_N(p, q)\prod_{n=1}^{N-1} \times P_n\left(p\prod_{l=n+1}^{N} M_l, q\prod_{l=n+1}^{N} M_l\right); \quad (20)$$

$$M = \prod_{n=1}^{N} M_n. \quad (21)$$

Equations (20) and (21) are most readily obtained by generalizing Eqs. (16) and (17) for the bipupil case.

It is worth noting that the considerations leading to Eqs. (14) to (18) have not required any restrictions upon the object function $f(x_0, y_0, p_0, q_0)$. In general, the problems of phase microscopy become too difficult unless in fact or as a first approximation $f(x_0, y_0, p_0, q_0)$ is independent of p_0, q_0, the optical direction cosines of the incident bundles of rays.

An Achromatic Ultraviolet Microscope Objective

L. V. FOSTER AND E. M. THIEL
Scientific Bureau, Bausch & Lomb Optical Co., Rochester, New York

(Received March 1, 1948)

Microscopes have been used with monochromatic illumination because achromatism in the ultraviolet has not been common. The authors have designed and made a few different objectives achromatized for a band width of about 100A in the region 2700A. These objectives make use of fused quartz and fluorite. They can be used in other regions of the ultraviolet spectrum, but require readjustment of focus.

MICROSCOPE objectives for use in the ultraviolet region were developed by Köhler and von Rohr. They used two materials, fused quartz having an ^nD of 1.4585 and fluorite having an ^nD of 1.4338, to secure correction for spherical aberration at 2750A.[1] In the early 1900's Dr. von Rohr[2] designed a series of objectives made of a single material, fused quartz, and corrected them for spherical aberration and sine condition. These objectives consisted of a series of aplanatic meniscus lenses and could, therefore, be corrected for only one wave-length. More recently Zeiss introduced a series of all-quartz monochromats corrected for 2750A.[3]

Some time ago the writers felt that one could correct microscope objectives for chromatic aberration over a short region in the ultraviolet spectrum. This seemed possible since the dispersions of fused quartz and fluorite are unlike, fused quartz having a ν value of 68 and fluorite having a ν value of 95.

A 2.5 mm fused quartz objective consisting of a hemisphere and six aplanatic meniscus lenses calculated to give minimum spherical aberration at 2750A had 4-mm axial chromatic aberration in a band width of 10A. This objective had 0.85 numerical aperture, for which the Rayleigh limit for chromatic aberration was 1.7 mm or somewhat less than half the actual aberration.

The first step in exploring the possibilities of correcting chromatic aberration was to replace the last three meniscus lenses with one fused quartz fluorite achromatic doublet. The chromatic aberration over the same 10A was reduced to 2.8 mm. Since this chromatic aberration approached the Rayleigh limit, and it was desirable to know how much chromatic aberration remained over a broader ultraviolet band, the axial image distance for 2650A was calculated. For this 100A band the aberration was 28 mm. A second fused quartz fluorite achromat was added and the chromatic aberration was reduced to 19 mm. A third fused quartz fluorite achromat further reduced it to 11 mm. This aberration was still much too great and the crown (fluorite) elements were made stronger to further reduce the chromatic aberration to 7.0 mm. At this point in the design the spherical aberration and sine condition were tested and found to be excessive. It must be understood that through the first four elements this system was aplanatic but not achromatic. The lenses beyond the hemisphere forming such a microscope objective had very little dioptric power. An aplanatic

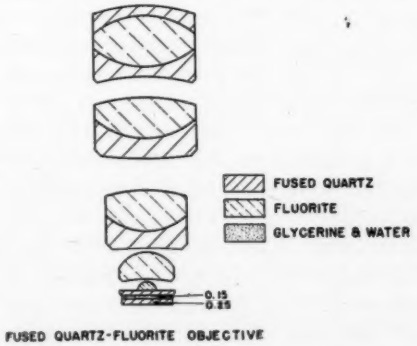


FIG. 1.

[1] A. Köhler, "Mikrophotographische Untersuchungen mit ultraviolettem Licht," Zeits. f. wiss. Mikroskopie u. f. mikroskopische Technik, Band XXI, 1904, p. 129–165 u 273–304.

[2] A. Köhler, "A Microscopic Arrangement for Ultra-Violet Light and Investigations of Organic Substances, With Its Use." The Optical Instrument Monthly, June, 1905.

[3] Zeiss-Objektive und Okulare für Mikroskopie und Mikrophotographie. Verschiedene Hilfsapparate und Zubehör. (Carl Zeiss Catalog) Mikro 367, 1940.

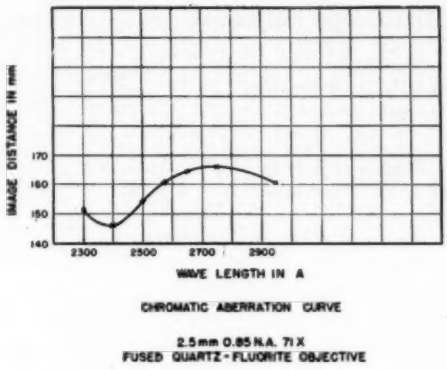


FIG. 2.

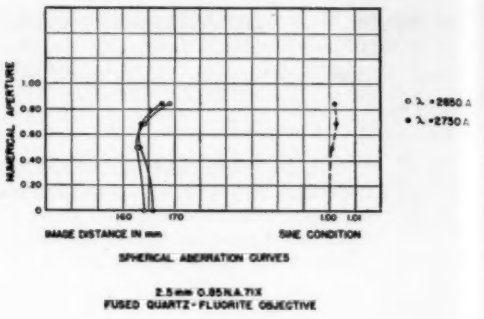


FIG. 3.

meniscus lens can have a magnification equal only to its index of refraction. If too many aplanatic meniscus lenses are introduced beyond the hemisphere before achromatic doublets are brought into play, achromatism cannot be complete. It was necessary to depart from the aplanatic form in order to crowd chromatic correction near the front of the objective. Since fluorite has a much lower dispersion than fused quartz all positive elements were made of fluorite and the final design consisted of a fluorite hemisphere cemented to a fused quartz plate, a fluorite meniscus, two fused quartz fluorite doublets, and a fused quartz fluorite triplet, as shown in Fig. 1. This objective had 1.5 mm chromatic aberration between 2650A and 2750A, the Rayleigh limit being 1.95 mm. Chromatic aberration was calculated for shorter and longer wave-lengths and is shown plotted in Fig. 2. The spherical aberration for 2750A was held to 1.5 mm. Fig. 3 shows the spherical aberration curves for both 2650A and 2750A. It can be seen from these curves that the spherical zone is only about 2.4 mm undercorrected. The Rayleigh limit for spherical aberration is 7.8 mm and for the zonal aberration is 11.2 mm. On Fig. 3, in dashed lines, is plotted the sine condition. Perfect sine condition is represented by a straight line.

Achromatic objectives for use in the ultraviolet region are in constant demand by the biophysicist. They are used to increase resolving power and to disclose characteristics of material selective in absorption at different wave-lengths in the ultraviolet. For instance, 10 percent nucleic acid at 2600A in a thickness of 5 mu has an absorption of 90 percent, while 10 percent protein under the same circumstances has an absorption of 2 percent.[4] At 2800A the same quantity of protein has 5 percent absorption. It is desirable to observe biological material in the ultraviolet region where differential absorption is pronounced. Dr. Caspersson of the Karolinska Institute in Stockholm, Sweden, has observed these differences with the Köhler fused quartz monochromats using a cadmium spark for illumination at 2750A and a magnesium spark for illumination at 2800A. His observations have been made with a Köhler fluorescent eyepiece. The illumination of the specimen in ultraviolet light offered some problems. In order to isolate a part of the spectrum it was necessary either to

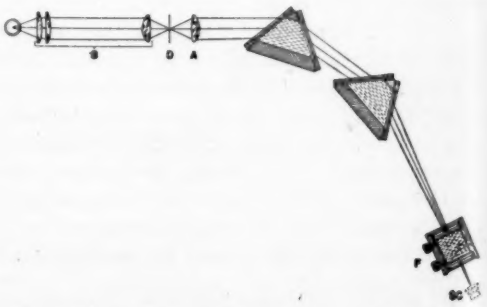


FIG. 4. Monochromator: The light source is imaged at *D* by the condenser lenses *B*. *D* is a diaphragm which is imaged by lens *A* through the 70 degree water prisms and the Bäckstrom filter *F* in the aperture of the substage condenser *SC*. All optical parts are either crystal quartz or fused quartz.

[4] T. Caspersson, "Über den chemischen Aufbau der Strukturen des Zellkernes"—April, 1936.

FIG. 5. Spectrum of *U*-Viarc. Figure 6 was taken using the portion from 2650A to and including 2750A.

use filters or a monochromator. Since filters transmitting narrow regions in the ultraviolet do not exist it became necessary to use a monochromator. After having developed the system of illumination it was found that Köhler had described in 1899 a similar monochromator used only for the visual spectrum.[5] The present monochromator consisted of two 70° water prisms with fused quartz windows, a quartz lens *A* and a diaphragm (*D*), as shown in Fig. 4. The source of light was a *U*-Viarc which was imaged on the diaphragm with quartz lenses *B*. It was desirous to illuminate all parts of the aperture of the objective with light from 2650A to 2750A. This meant that it would be necessary to spread this band across the free aperture of the 0.85 N. A. fused quartz substage condenser (*SC*). This fixed the distance from the second prism to the substage condenser at about 14 inches. In most monochromators the entrance and exit apertures are slits. The slit was replaced with a round diaphragm (*D*) at the entrance aperture of the monochromator. The quartz lens *A* magnified this aperture diaphragm 20× to fill the 10-mm aperture of the substage condenser. In order evenly to illuminate the field of the microscope with light the quartz lens *A* must be evenly illuminated, since it is imaged in the field by the substage condenser. It also became the field limiting aperture. Lens system *B* is used to image the source on the diaphragm and to fill lens *A* with light. In any monochromator, stray light will cause trouble. The stray light from strong mercury lines outside the 2650–2750A region was removed by using directly in front of the substage condenser a liquid filter (*F*) described by Bäckstrom,[6] called to the writers' attention by Dr. George I. Lavin of the Rockefeller Institute for Medical Research. The filter solution was made by dissolving 46 grams of $NiSO_4+6H_2O$ and 14 grams of $CoSO_4+7H_2O$ in 70 ml. of distilled water. After the salts were dissolved the solution was made to 100 ml. This solution was placed in a 3 cm filter cell with quartz windows.

FIG. 6. Photomicrograph of lines ruled in silver made with the objective of Fig. 1 and a 7.5× quartz eyepiece at a magnification of 500×.

Test photographs were made with the 2.5 mm achromatic objective and a 7.5× quartz eyepiece. The test object used was made similar to an Abbe test plate. Silver was deposited on a slide made from Corning 791 high silica glass. A ground and polished fused quartz cover slip was cemented over the rulings. Since there was no convenient source of illumination with a continuous spectrum in the ultraviolet, the *U*-Viarc was chosen which has strong lines or bands of energy in this part of the spectrum, as shown in Fig. 5. There is a triplet at 2653A, a strong line at 2700A and two lines at about 2753A. This group of lines was used to illuminate the object. The result is shown in Fig. 6. The mercury spectrum has also a strong band at

[5] A. Köhler, "Beleuchtungsapparat für gleichmässige Beleuchtung mikroskopischer Objecte mit beliebigem einfarbigem Licht," Zeits. f. wiss. Mikroskopie, Band XVI, 1899, p. 1–29.

[6] H. L. J. Bäckstrom, Acta Radiologica, Vol. XXI, p. 327–334 (1940).

2537A. Pictures of the same object were made using the group of lines 2537A to and including 2653A by refocusing. These pictures were as good as Fig. 6. There is also a band at 2804A and 3024A with lines at 2897A, 2925A, and 2967A between them. Other test pictures made with illumination from 2804A to 3024A were also as good as Fig. 6. The objective is, therefore, usable in a 100A band in the region from 2500A to 3000A. Possibly a wider band of illumination could be used. It has never been tested for more than the 100A band width.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 8 AUGUST, 1948

Calibration of a Photo-Multiplier Photometer*

J. R. DEVORE
Research Division, Technical Department, The New Jersey Zinc Company (of Pa.), Palmerton, Pennsylvania
(Received April 1, 1948)

A photoelectric photometer which has been constructed for the measurement of brightness of luminescent materials has been calibrated so that the brightness values will correlate with direct visual observations. The calibration curves are presented and are shown to be in agreement with published eye-sensitivity data.

A PHOTOELECTRIC photometer has been constructed for the purpose of measuring the brightness of luminescent pigments and paints. The instrument is used for making the brightness measurements involved in the routine control tests on commercial production of luminescent pigments where only a limited number of different colors are encountered. The photometer uses an RCA 1P21 photo-multiplier tube which is specially selected for work at low levels of brightness. The photo-multiplier tube is operated on a "B" battery power supply at 90 volts per stage. A spotlight galvanometer with an Ayrton shunt is the indicating meter for the instrument.

Figure 1 is a photograph of the apparatus. The photo-multiplier tube is mounted within the brass tube on the end of the cabinet that contains the batteries and switches. This photometer is designed for the measurement of both fluorescent and phosphorescent brightnesses. When measuring fluorescent brightness, it is necessary to filter out the ultraviolet light reflected from the sample. This is accomplished by removing the end cap of the brass tube and inserting a sleeve liner having a "Noviol" filter (Corning No. 3389) mounted over the aperture.

FIG. 1. Photoelectric low brightness photometer.

In order to be used for both phosphorescent and fluorescent measurements, the photometer must measure brightnesses ranging from 0.01 effective microlamberts to something over 100 effective microlamberts. An effective microlambert is a unit of brightness based on the assumption of a linear relationship between brightness and intensity for a total radiator having a color temperature of 2360°K. The fundamental measurement of brightness (in effective microlamberts) of a given source is carried out by heterochromatic matching with a visual photometer having a field lighted at a color temperature of 2360°K.

The full range of phosphorescent brightness is handled by the use of the shunt. However, when the brightness exceeds 10 microlamberts, the current in the photo-multiplier approaches a saturation value so the increase in current is no longer proportional to the increase in brightness. Consequently, in order to cover the range of

* Presented at the meeting of the Optical Society of America in Cincinnati, Ohio, October 23–25, 1947.

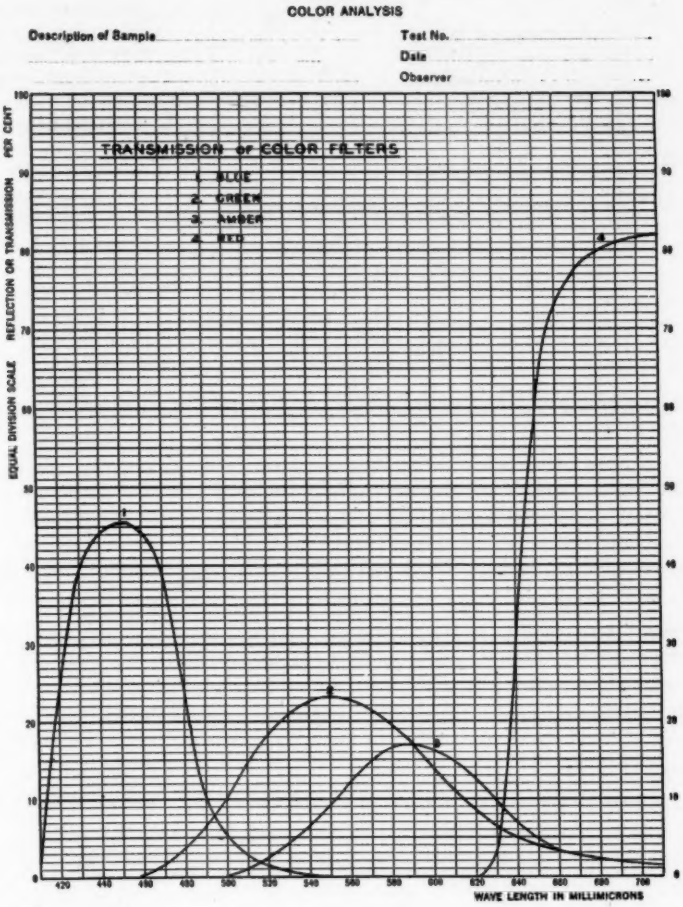


FIG. 2. Transmission of color filters.

brightnesses involved in fluorescent measurements, the sleeve liner carrying the "Noviol" filter has a restricting aperture so that the photo-currents will fall in the proportionate range.

It is desired to have the measurements made with this photometer correlate with brightnesses observed when the samples are viewed with the eye. The spectral sensitivity and absolute sensitivity of the eye changes as the brightness decreases so the photo-current will not be a linear function of the brightness. Since this change in eye sensitivity is a gradual shift rather than a sharp transition, it is not possible to use a filter to correct the spectral response of the cell. Instead, it is necessary to determine calibration curves for the various colors. An excellent discussion of the change in eye sensitivity and its

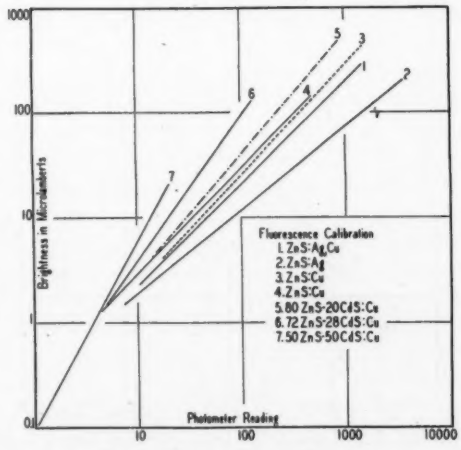


FIG. 3. Fluorescence calibration.

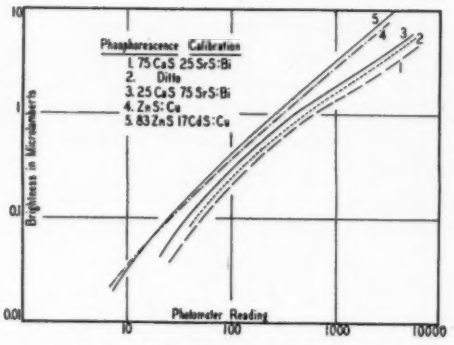


FIG. 4. Phosphorescence calibration.

attendant problems is contained in the paper by R. P. Teele.[1]

In order to have the photoelectric photometer measurements correlate with brightnesses by visual observation, the calibration was carried out by measuring samples and light sources both with the photoelectric instrument and with a low brightness visual photometer. The low brightness visual photometer matches the samples against a 2360°K field so that the measured brightnesses do correlate with direct visual observation.

Calibrations were carried out in three different ways:

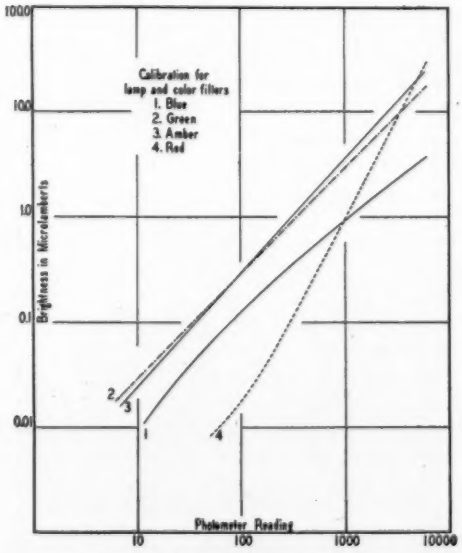


FIG. 5. Calibration for lamp and color filters.

[1] R. P. Teele, J. Research Nat. Bur. Stand. **34**, 325 (1945).

1. Panels covered with fluorescent paints of different colors were measured with the visual photometer with fixed conditions of excitation and then measured with the photoelectric photometer under the same excitation.

2. Panels covered with phosphorescent paints of different colors were excited to high brightness and then the decay curves were determined concurrently with the two instruments. The visual photometer yielded decay curves of brightness against time while the photoelectric photometer yielded decay curves of photometer reading against time. From corresponding pairs of curves the photometer calibration curves of photometer reading against brightness are obtained.

3. A light box was built having an opal screen illuminated by a Mazda lamp so as to have a color temperature of 2360°K. This light box was so designed that the intensity of light could be varied by inserting diaphragms between the lamp and the screen. A colored glass filter inserted between the lamp and screen produced a colored field which then was measured with the photoelectric photometer and with the visual photometer.

A set of four color filters having red, amber, green, and blue transmissions was used for this work. The transmission curves for these filters are shown in Fig. 2. These filters produced such strongly colored fields that measurements with the visual photometer, which involves heterochromatic matching, were difficult. Consequently the measurements were made by several individuals and a weighted average of the measurements was taken as the brightness value for each condition of measurement.

In all of this work the unit of brightness is the effective microlambert. In all the calibration curves, the value designated as photometer reading is the product of the galvanometer reading and the shunt factor.

1. The calibration curves for the fluorescence measurements are shown in Fig. 3. Sample 1 is a bluish white color. Sample 2 is blue, 3 and 4 are green, 5 is yellow, 6 is orange, and 7 is red orange. It will be seen that the calibrations are straight lines when plotted on a log-log scale, and that the slope consistently increases as the color moves through the spectrum from blue to red. It is just by chance that in this brightness range, the rate of change of eye sensitivity is such as to yield straight lines on a log-log plot.

By the definition of an effective microlambert, measurements made on a 2360°K lamp would yield a straight line on a linear scale which would be a 45° line on a log-log scale. It would be expected that the closer the color of a sample matched the 2360°K color, the closer the calibra-

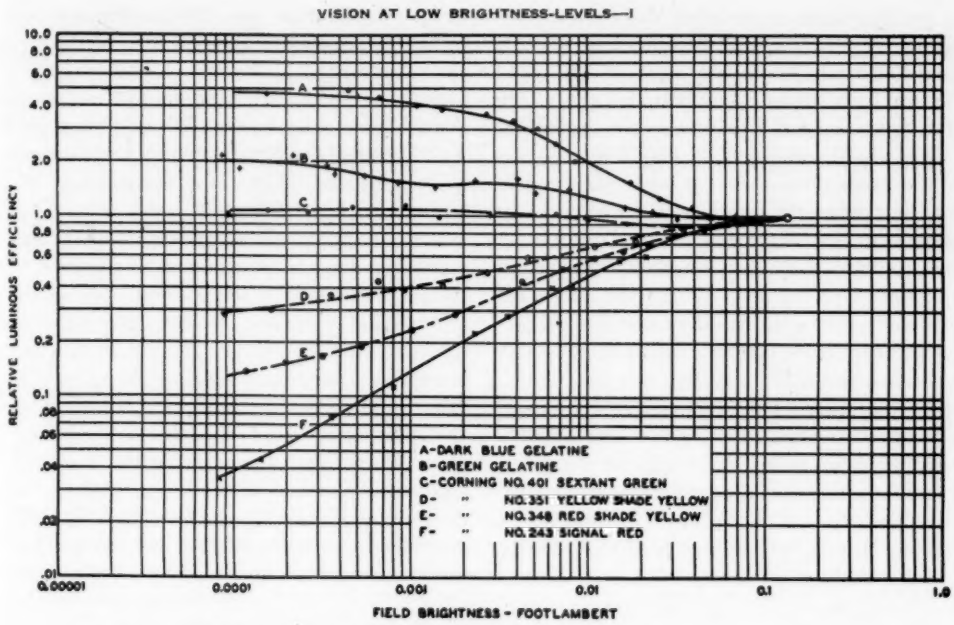


FIG. 6. Effect of field brightness on the relative luminous efficiency of energy from a tungsten lamp at color-temperature 2848 degrees Kelvin transmitted by filters *A* to *F*.

tion curve would be to 45°. However, the curve for the yellow sample (No. 5) which is slightly more green than 2360°K is considerably steeper than 45° and the sample coming closest to 45° is the green (No. 4). This point will be discussed later.

2. Considering, now, the calibration for phosphorescent samples, the calibration curves are shown in Fig. 4. On this graph, curves 1 and 2 are for blue pigments of slightly different shades. Sample 2 is slightly more green than Sample 1. Curve 3 is for a greenish blue pigment. Curve 4 is a green sample, and curve 5 is a yellow pigment. The differences in color among the first three samples are not great and that is shown by the fact that the curves are close together and similar in character.

At the high brightness end, these curves have the same appearance as the fluorescence curves; that is, they are nearly straight, and the slope increases as the color shifts toward red. However, as the brightness decreases, the lines begin to curve and the curvature is greatest for the blue samples so that at low brightness the slope-color relationship does not hold. This effect is even more pronounced in the measurement of the colored lights.

For routine use of this instrument, two of these calibration curves (Nos. 2 and 5) are sufficient because the differences among the first three curves are no greater than the variations in visual measurements made on one sample by different observers, and the same is true for the two remaining curves.

3. The results of the measurements on the colored lights are shown in Fig. 5. Here again, as in the fluorescent samples, at high brightnesses the curves are approximately straight lines with the slope of the lines increasing as the color shifts from blue to red. As the brightness decreases the lines begin to curve. Because these filters produce more saturated colors than the phosphorescent paints, the curvature is more pronounced. The curvature shows a consistent trend with color; thus the blue line is markedly concave downward, the green is slightly concave downward, the amber is slightly concave upward, and the red is strongly concave upward. The green curve is closest to a 45° line, which is in agreement with the results observed in the fluorescent samples.

These various results agree well with eye sensitivity measurements published in the litera-

ture. In Fig. 6 are reproduced the curves published by A. H. Taylor[2] showing the relative luminous efficiency plotted against the field brightness for various colors. On this graph the 2360°K field would fall on a straight, horizontal line at a visual efficiency of 1.0. These curves show the same phenomena as were observed in the calibration of the photoelectric photometer; namely (1) at the brightness levels of fluorescent pigments (0.005 to 0.05 footlambert) the curves approximate straight lines of different slopes for the different colors, (2) at the very low brightnesses the curvature changes continuously from concave downward for the blue to concave upward for the red, and (3) green (Curve *C*) most closely approximates the line for the 2360°K field rather than yellow as would be expected. This last fact indicates that color-eye sensitivity relationships do not hold when comparing a continuous spectrum source such as the 2360°K field with a limited spectral band source such as the lamp and filter combinations. Consequently, the calibration of an instrument of this type should be carried out experimentally with materials of the type that will be measured thereafter. Calculation of the calibration curves from eye sensitivity data is susceptible to considerable error.

Spectrophotometric systems such as have been described recently by A. E. Hardy[3] and by Plymale[4] have the advantage that they do not require separate calibrations for samples of different colors. It is merely necessary to have a system (either mechanical or mathematical) for correcting the non-linear spectral response of the system. However, in the present application the information required is not the spectrophotometric data but rather the integrated output corrected for eye response so that the apparatus and calibration curves just described are more useful.

Reproducibility of measurements with this photoelectric instrument is better than the reproducibility of measurements with a visual instrument. Furthermore, the measurements can be made much more rapidly and easily with the photoelectric instrument. When the calibration curves, that have been shown, are used, the results fulfill the requirement that the measurements should correlate with visual observations and gradings.

[2] A. H. Taylor, Illum. Eng. **38**, 89 (1943).

[3] A. E. Hardy, Electrochem. Soc. Preprint 91–98.
[4] W. S. Plymale, Jr., J. Opt. Soc. Am. **37**, 399 (1947).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 8 AUGUST, 1948

I.C.I. Specifications of Difference Limens for Munsell Hue, Value, and Chroma

JOSEPHINE G. BRENNAN AND SIDNEY M. NEWHALL
Color Control Department, Eastman Kodak Company, Rochester 4, New York

(Received March 26, 1948)

In an earlier two-year study, color difference limens of three observers were determined separately for each of the Munsell analogs of hue, saturation, and lightness at each of a number of locations in the surface-color solid. The present paper makes these data available in terms of the standard I.C.I. system and so helps to make that system more visually meaningful. In addition to detailed tabulations of the x, y, and Y specifications, the means of the chromatic difference limens of the three observers are plotted in the I.C.I. chromaticity coordinates at three levels of reflectance. Absolute chroma limens for Munsell red, yellow, green, blue, and purple are tabulated at seven levels of reflectance. The general order of magnitude of all the liminal determinations may be indicated by the fact that the trichromatic coefficients and reflectance figures representing them are confined to the third and fourth decimal places. In terms of distances in the chromaticity diagram, the general mean hue and chroma limens were, respectively, 0.0023 and 0.0035; and the mean Fechner fraction from the value data turned out to be 0.0087.

A RATHER good indication of the order of magnitude and variation of threshold color differences in the Munsell system of notation was secured in 1942.[1] Difference limens for hue, value, and chroma were individually obtained by a technique which was described in detail, from

[1] B. R. Bellamy and S. M. Newhall, "Attributive limens in selected regions of the Munsell color solid," J. Opt. Soc. Am. **32**, 465 (1942).

TABLE I. Hue difference limens of 3 observers at Munsell value levels 2/, 5/, and 8/, as expressed in terms of I.C.I. Δx, Δy, ΔY.

	Origin			Obs. B			Obs. M			Obs. N			Mean		
Munsell	x	y	Y	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY
R 2/2	.3687	.3079	.0400	.0016	.0015	.0001	.0019	.0017	.0001	.0016	.0014	.0001	.0017	.0015	.0001
R 2/6	.4218	.2978	.0461	.0002	.0002	.0000	.0027	.0025	.0001	.0011	.0008	.0000	.0013	.0012	.0000
Y 2/2	.3527	.3611	.0472	.0000	.0000	.0000	.0003	.0005	.0001	.0003	.0005	.0001	.0002	.0003	.0001
G 2/2	.2806	.3489	.0373	.0033	.0017	.0005	.0011	.0009	.0002	.0018	.0006	.0002	.0021	.0011	.0003
B 2/2	.2550	.2944	.0381	.0009	.0037	.0000	.0004	.0014	.0000	.0004	.0014	.0000	.0006	.0022	.0000
P 2/2	.2966	.2467	.0351	.0012	.0003	.0000	.0007	.0000	.0000	.0012	.0001	.0000	.0010	.0001	.0000
P 2/6	.2820	.1909	.0346	.0017	.0002	.0001	.0017	.0002	.0001	.0020	.0000	.0001	.0018	.0001	.0001
	Mean			.0013	.0011	.0001	.0013	.0010	.0001	.0012	.0007	.0001	.0012	.0009	.0001
R 5/2	.3507	.3176	.1954	.0019	.0010	.0015	.0008	.0004	.0006	.0016	.0009	.0012	.0014	.0008	.0011
R 5/6	.4337	.3167	.1998	.0025	.0011	.0003	.0062	.0029	.0007	.0037	.0017	.0004	.0041	.0019	.0005
R 5/10	.4855	.3202	.2128	.0055	.0027	.0007	.0100	.0050	.0014	.0057	.0028	.0008	.0071	.0035	.0010
Y 5/2	.3446	.3547	.2117	.0007	.0011	.0013	.0007	.0011	.0013	.0008	.0013	.0014	.0007	.0012	.0013
Y 5/6	.4272	.4478	.1984	.0002	.0013	.0003	.0002	.0014	.0004	.0003	.0018	.0005	.0002	.0015	.0004
G 5/2	.2930	.3418	.1888	.0019	.0011	.0011	.0017	.0010	.0010	.0028	.0016	.0016	.0021	.0012	.0012
G 5/6	.2625	.3944	.1869	.0027	.0015	.0006	.0040	.0020	.0008	.0044	.0022	.0009	.0037	.0019	.0008
B 5/2	.2692	.2951	.1917	.0008	.0025	.0014	.0011	.0032	.0018	.0008	.0023	.0013	.0009	.0027	.0015
B 5/6	.2106	.2630	.1940	.0017	.0057	.0016	.0020	.0066	.0018	.0017	.0056	.0015	.0018	.0060	.0016
P 5/2	.3023	.2861	.1959	.0013	.0002	.0006	.0013	.0002	.0006	.0009	.0002	.0004	.0012	.0002	.0005
P 5/6	.2921	.2424	.1965	.0008	.0001	.0004	.0013	.0002	.0007	.0015	.0002	.0008	.0012	.0002	.0006
P 5/10	.2820	.2077	.2039	.0017	.0005	.0006	.0018	.0005	.0006	.0023	.0007	.0008	.0019	.0006	.0007
	Mean			.0018	.0016	.0009	.0026	.0020	.0010	.0022	.0018	.0010	.0022	.0018	.0009
R 8/2	.3260	.3167	.6021	.0000	.0002	.0001	.0002	.0008	.0004	.0001	.0004	.0002	.0001	.0005	.0002
Y 8/2	.3337	.3426	.5570	.0000	.0005	.0009	.0001	.0009	.0019	.0000	.0007	.0014	.0000	.0007	.0014
Y 8/6	.3902	.4074	.5784	.0000	.0003	.0003	.0000	.0014	.0015	.0000	.0012	.0014	.0000	.0010	.0011
Y 8/8	.4177	.4429	.5732	.0028	.0016	.0020	.0031	.0017	.0021	.0025	.0013	.0017	.0028	.0015	.0019
G 8/2	.3017	.3385	.5694	.0007	.0002	.0005	.0014	.0005	.0011	.0011	.0004	.0009	.0011	.0004	.0008
G 8/6	.2844	.3552	.5595	.0022	.0022	.0011	.0038	.0037	.0020	.0028	.0028	.0015	.0029	.0029	.0015
B 8/2	.2881	.3105	.6179	.0008	.0017	.0010	.0006	.0014	.0009	.0005	.0009	.0006	.0006	.0013	.0008
P 8/2	.3055	.3007	.5897	.0007	.0002	.0016	.0012	.0003	.0030	.0009	.0002	.0022	.0009	.0002	.0023
	Mean			.0009	.0008	.0009	.0013	.0013	.0016	.0010	.0010	.0012	.0012	.0011	.0012

three experienced observers at various representative locations in the Munsell solid. Since these data appear to be still the only comparable liminal measurements of the recognized analogs of hue, saturation, and lightness, it has seemed worth while to make them available in terms of the standard I.C.I. system of color specification, and that is the purpose of this paper. To that end, then, 369 individual limens were converted by conventional methods to the I.C.I. notation, that is, x, y, and Y relative to Illuminant C.[2] These I.C.I. computations were made under the direction of Robert H. Morris.* The breakdown by observers has been included in the following tables to reveal the order of magnitude of the indivudual differences.

[2] D. Nickerson, *Color Measurement* (U. S. Dept. Agri., 1946), Misc. Publ. **580**, p. 22.

* Though they are not responsible for its contents, this paper has the benefit of comments and criticisms by the above-mentioned, as well as by Carl F. J. Overhage and David L. MacAdam; and these helpful suggestions are herewith acknowledged.

TABLES AND RESULTS

The first three tables are identical in arrangement, the only difference being that Table I presents the hue difference limens, Table II the chroma difference limens, and Table III the value difference limens. The limens in each table were taken at value levels 2/, 5/, and 8/, and at the same 27 locations or origins distributed through Munsell color space. The specifications of these origins are given in both Munsell and I.C.I. notation in the left-hand columns of each table. In the columns to the right are given the corresponding limens in terms of I.C.I. Δx, Δy, and ΔY, for each of the three observers B, M, and N separately, and the mean of all three. Consider, for the sake of illustration, the first entry for observer B in each of these tables. At the location in the Munsell color solid specified by either R 2/2, or $x=0.3687$, $y=0.3079$, and $Y=0.0400$, Table I shows (in the first row) the hue limen for observer B to be $\Delta x=0.0016$,

$\Delta y = 0.0015$, and $\Delta Y = 0.0001$; Table II shows the corresponding chroma limen to be $\Delta x = 0.0010$, $\Delta y = 0.0001$, and $\Delta Y = 0.0003$; and Table III shows the corresponding value limen to be $\Delta x = 0.0002$, $\Delta y = 0.0001$, and $\Delta Y = 0.0002$. The delta-symbol is used to signify merely that each individual figure is a difference between two trichromatic coefficients or Y specifications. All of these differences were treated as absolute quantities, and the means of the tables are not zero.

In further explanation of the data it should be remarked that the hue and chroma limens in Tables I and II are represented chiefly by the differences Δx and Δy, because each limen was taken at a given level of luminous reflectance. The fact that ΔY sometimes assumes relatively large figures in these tables can be ascribed to technical error because value is defined in terms of percent luminous reflectance. The value limens in Table III are represented chiefly by the differences ΔY. This fact is apparent from the generally small Δx and Δy figures together with the systematic, to-be-expected, changes in the ΔY figures, as a function of value, through the three sections of the table.

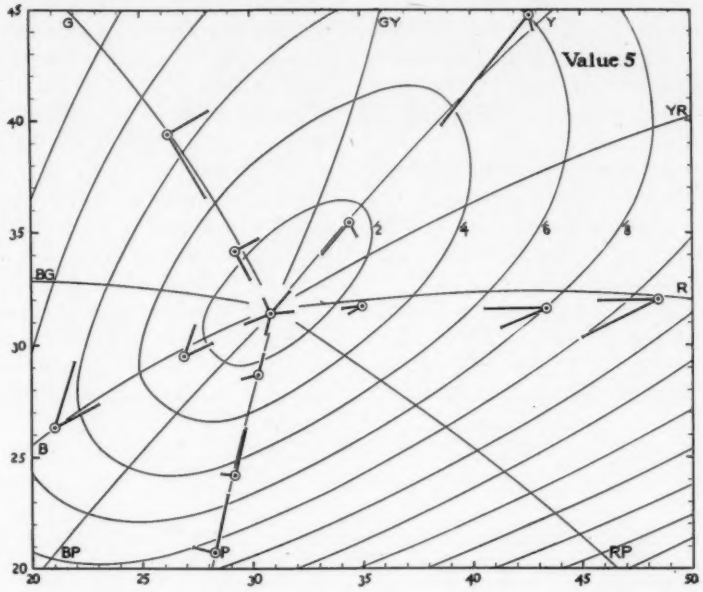
The fourth table is arranged like the first three tables, but whereas they contain difference limens only, Table IV contains absolute limens for chroma as well as achromatic difference limens for value. All of these limens were taken from origins on the central neutral axis of the color solid with replication at each of the value levels 2/, 3/, 4/, 5/, 6/, 7/, and 8/. These value levels are approximately equivalent to the respective subcommittee reflectance figures for I.C.I. Y, *viz.*, 0.0312, 0.0655, 0.1200, 0.1977, 0.3005, 0.4306, and 0.5910. Absolute chroma limens for Munsell 5*R*, 5*Y*, 5*G*, 5*B*, and 5*P* will be found in the first five horizontal sections of the table which are headed Red, Yellow,

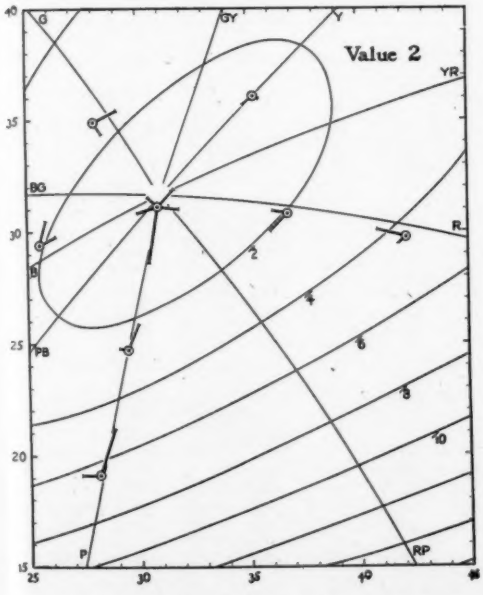
TABLE II. Chroma difference limens of 3 observers at Munsell value levels 2/, 5/, and 8/, as expressed in terms of I.C.I. Δx, Δy, ΔY.

Munsell	Origin			Obs. B			Obs. M			Obs. N			Mean		
	x	y	Y	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY
R 2/2	.3687	.3079	.0400	.0010	.0001	.0003	.0015	.0003	.0003	.0015	.0003	.0003	.0013	.0002	.0003
R 2/6	.4218	.2978	.0461	.0024	.0004	.0006	.0028	.0006	.0006	.0026	.0004	.0005	.0026	.0005	.0006
Y 2/2	.3527	.3611	.0472	.0007	.0006	.0003	.0009	.0008	.0004	.0013	.0013	.0006	.0010	.0009	.0004
G 2/2	.2806	.3489	.0373	.0006	.0003	.0001	.0008	.0009	.0002	.0011	.0015	.0003	.0008	.0009	.0002
B 2/2	.2550	.2944	.0381	.0014	.0002	.0003	.0018	.0007	.0003	.0017	.0006	.0004	.0016	.0005	.0003
P 2/2	.2966	.2467	.0351	.0008	.0049	.0005	.0004	.0014	.0001	.0005	.0013	.0002	.0006	.0025	.0003
P 2/6	.2820	.1909	.0346	.0014	.0061	.0004	.0004	.0024	.0001	.0010	.0037	.0003	.0009	.0041	.0003
	Mean			.0012	.0018	.0004	.0012	.0010	.0003	.0014	.0013	.0004	.0013	.0014	.0003
R 5/2	.3507	.3176	.1954	.0017	.0002	.0002	.0018	.0002	.0002	.0017	.0003	.0002	.0017	.0002	.0002
R 5/6	.4337	.3167	.1998	.0040	.0000	.0003	.0069	.0001	.0005	.0059	.0001	.0005	.0056	.0001	.0004
R 5/10	.4855	.3202	.2128	.0041	.0001	.0006	.0078	.0003	.0011	.0046	.0001	.0006	.0055	.0002	.0008
Y 5/2	.3446	.3547	.2117	.0023	.0027	.0014	.0026	.0030	.0015	.0022	.0026	.0013	.0024	.0028	.0014
Y 5/6	.4272	.4478	.1984	.0065	.0073	.0003	.0088	.0099	.0004	.0088	.0129	.0004	.0080	.0100	.0004
G 5/2	.2930	.3418	.1888	.0013	.0024	.0001	.0011	.0020	.0001	.0014	.0027	.0002	.0013	.0024	.0001
G 5/6	.2625	.3944	.1869	.0025	.0043	.0002	.0036	.0063	.0002	.0038	.0066	.0002	.0033	.0057	.0002
B 5/2	.2692	.2951	.1917	.0012	.0006	.0000	.0030	.0014	.0001	.0035	.0016	.0001	.0026	.0012	.0001
B 5/6	.2106	.2630	.1940	.0036	.0019	.0002	.0034	.0019	.0001	.0051	.0028	.0002	.0040	.0022	.0002
P 5/2	.3023	.2861	.1959	.0003	.0017	.0003	.0005	.0020	.0004	.0005	.0020	.0004	.0004	.0019	.0004
P 5/6	.2921	.2424	.1965	.0011	.0046	.0005	.0010	.0042	.0005	.0008	.0038	.0004	.0010	.0042	.0005
P 5/10	.2820	.2077	.2039	.0012	.0049	.0010	.0011	.0046	.0009	.0014	.0057	.0011	.0012	.0051	.0010
	Mean			.0025	.0026	.0004	.0035	.0030	.0005	.0033	.0034	.0005	.0031	.0030	.0005
R 8/2	.3260	.3167	.6021	.0007	.0003	.0017	.0009	.0004	.0022	.0008	.0004	.0020	.0008	.0004	.0020
Y 8/2	.3337	.3426	.5570	.0007	.0009	.0006	.0016	.0017	.0014	.0009	.0010	.0008	.0011	.0012	.0009
Y 8/6	.3902	.4074	.5784	.0022	.0024	.0001	.0049	.0054	.0002	.0024	.0025	.0001	.0032	.0034	.0001
Y 8/8	.4177	.4429	.5732	.0054	.0062	.0001	.0072	.0083	.0001	.0040	.0046	.0001	.0055	.0064	.0001
G 8/2	.3017	.3385	.5694	.0005	.0006	.0002	.0010	.0014	.0005	.0006	.0008	.0003	.0007	.0009	.0003
G 8/6	.2844	.3552	.5595	.0018	.0021	.0009	.0020	.0022	.0009	.0022	.0025	.0011	.0020	.0023	.0010
B 8/2	.2881	.3105	.6179	.0009	.0004	.0015	.0015	.0006	.0027	.0009	.0004	.0016	.0011	.0005	.0019
P 8/2	.3055	.3007	.5897	.0005	.0014	.0010	.0007	.0018	.0012	.0005	.0014	.0010	.0006	.0015	.0011
	Mean			.0016	.0018	.0008	.0025	.0027	.0011	.0015	.0017	.0009	.0019	.0021	.0009

Green, Blue, Purple, respectively; and the differential achromatic value limens will be found in the following or final section which is headed Value.

Tables I to IV have something of a reference function. Thus if one happens to be interested in the I.C.I. limen at or near any of the various





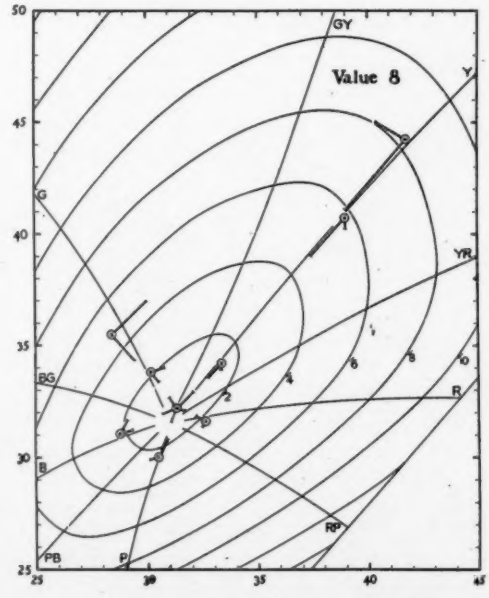


FIG. 1. Average limens (times 5) of Munsell hue and chroma at value levels 2/, 5/, and 8/, as plotted in I.C.I. x, y diagrams showing the O.S.A. subcommittee contours of hue and chroma. An origin color is represented by the center of each small circle, and each of the two or five short lines radiating therefrom represents an average limen (times 5) of three experienced observers

locations listed in the left-hand columns, he can read in the columns to the right the individual and mean liminal figures which were obtained from the three experienced observers. The value of this reference function is necessarily limited by the generally high sensitivity of limens to experimental conditions, by the specific directions in which the limens were taken, by the variability of the obtained data, and by the fact that the I.C.I. color gamut is so far from being covered. Somewhat related data are available in the form of I.C.I. specifications of near-threshold chromaticities which do represent nearly all the regions of the chromaticity diagram.[3,4]

Possible misunderstanding may be avoided by noting explicitly that the obtained figures for hue and chroma do not indicate directly the lengths of the lines in the I.C.I. chromaticity surface corresponding to the various difference limens. The obtained figures are simply the differences in the x and y trichromatic coefficients of the standard specification. The I.C.I. linear distances run somewhat larger than Δx and Δy, as evident from the usual formula, $(\overline{\Delta x^2}+\overline{\Delta y^2})^{\frac{1}{2}}$.

Figure 1 illustrates this plotting of limens by lines in I.C.I. chromaticity diagrams which include for comparison the hue and chroma contours recommended by the O.S.A. subcommittee in 1943.[5] Each short straight line represents the average of the limens of three observers multiplied by 5. The centers of the small circles represent the origins from which the various limens were taken. These diagrams not only show how the hue and chroma limens look as plotted in the I.C.I. chromaticity plane but also show the

TABLE III. Value difference limens of 3 observers at Munsell value levels 2/, 5/, and 8/, as expressed in terms of I.C.I. Δx, Δy, ΔY.

Munsell	Origin			Obs. B			Obs. M			Obs. N			Mean		
	x	y	Y	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY
R 2/2	.3687	.3079	.0400	.0002	.0001	.0002	.0000	.0000	.0001	.0003	.0003	.0004	.0002	.0001	.0002
R 2/6	.4218	.2978	.0461	.0001	.0001	.0001	.0001	.0005	.0005	.0006	.0003	.0004	.0003	.0003	.0003
Y 2/2	.3527	.3611	.0472	.0001	.0001	.0001	.0002	.0002	.0004	.0004	.0004	.0003	.0002	.0002	.0003
G 2/2	.2806	.3489	.0373	.0002	.0000	.0001	.0002	.0000	.0001	.0001	.0004	.0003	.0002	.0001	.0002
B 2/2	.2550	.2944	.0381	.0000	.0001	.0001	.0002	.0003	.0004	.0000	.0001	.0003	.0001	.0002	.0003
P 2/2	.2966	.2467	.0351	.0000	.0001	.0001	.0002	.0000	.0001	.0000	.0001	.0000	.0001	.0001	.0001
P 2/6	.2820	.1909	.0346	.0001	.0001	.0002	.0001	.0006	.0005	.0000	.0002	.0001	.0001	.0003	.0003
	Mean			.0001	.0001	.0001	.0001	.0002	.0003	.0002	.0003	.0003	.0002	.0002	.0002
R 5/2	.3507	.3176	.1954	.0000	.0000	.0018	.0002	.0000	.0025	.0002	.0001	.0022	.0001	.0000	.0022
R 5/6	.4337	.3167	.1998	.0001	.0000	.0021	.0002	.0000	.0025	.0001	.0000	.0021	.0001	.0000	.0022
R 5/10	.4855	.3202	.2128	.0005	.0001	.0031	.0007	.0003	.0044	.0005	.0002	.0034	.0006	.0002	.0036
Y 5/2	.3446	.3547	.2117	.0002	.0001	.0038	.0000	.0000	.0024	.0000	.0001	.0023	.0001	.0001	.0028
Y 5/6	.4272	.4478	.1984	.0000	.0001	.0003	.0000	.0001	.0021	.0001	.0001	.0007	.0000	.0001	.0010
G 5/2	.2930	.3418	.1888	.0002	.0002	.0025	.0001	.0001	.0019	.0001	.0001	.0032	.0001	.0001	.0025
G 5/6	.2625	.3944	.1869	.0001	.0001	.0003	.0001	.0001	.0009	.0000	.0002	.0008	.0001	.0001	.0007
B 5/2	.2692	.2951	.1917	.0001	.0001	.0015	.0002	.0000	.0023	.0001	.0000	.0020	.0001	.0000	.0019
B 5/6	.2106	.2630	.1940	.0002	.0003	.0015	.0002	.0002	.0025	.0002	.0002	.0021	.0002	.0002	.0020
P 5/2	.3023	.2861	.1959	.0000	.0001	.0026	.0000	.0001	.0031	.0000	.0002	.0026	.0000	.0001	.0028
P 5/6	.2921	.2424	.1965	.0000	.0002	.0022	.0001	.0002	.0025	.0001	.0002	.0025	.0001	.0002	.0024
P 5/10	.2820	.2077	.2039	.0001	.0001	.0022	.0001	.0000	.0025	.0001	.0001	.0027	.0001	.0001	.0025
	Mean			.0001	.0001	.0020	.0002	.0001	.0025	.0001	.0001	.0022	.0001	.0001	.0022
R 8/2	.3260	.3167	.6021	.0002	.0000	.0042	.0003	.0000	.0070	.0003	.0000	.0054	.0003	.0000	.0055
Y 8/2	.3337	.3426	.5570	.0001	.0000	.0037	.0001	.0000	.0050	.0001	.0000	.0040	.0001	.0000	.0042
Y 8/6	.3902	.4074	.5784	.0004	.0004	.0046	.0005	.0005	.0062	.0004	.0003	.0041	.0004	.0004	.0050
Y 8/8	.4177	.4429	.5732	.0006	.0006	.0047	.0006	.0006	.0050	.0006	.0006	.0048	.0006	.0006	.0048
G 8/2	.3017	.3385	.5694	.0001	.0000	.0026	.0001	.0001	.0050	.0001	.0001	.0034	.0001	.0001	.0037
G 8/6	.2844	.3552	.5595	.0003	.0003	.0046	.0003	.0002	.0038	.0002	.0002	.0027	.0003	.0002	.0037
B 8/2	.2881	.3105	.6179	.0000	.0000	.0041	.0000	.0001	.0056	.0000	.0001	.0053	.0000	.0001	.0050
P 8/2	.3055	.3007	.5897	.0000	.0002	.0052	.0000	.0001	.0067	.0000	.0000	.0041	.0000	.0001	.0053
	Mean			.0002	.0002	.0042	.0002	.0002	.0055	.0002	.0002	.0042	.0002	.0002	.0046

[3] D. L. MacAdam, "Visual sensitivities to color differences in daylight," J. Opt. Soc. Am. **32**, 247 (1942).
[4] W. D. Wright, *Researches on Normal and Defective Colour Vision* (Mosby, St. Louis, Missouri, 1947).
[5] S. M. Newhall, D. Nickerson, and D. B. Judd, "Final report of the O.S.A. subcommittee on the spacing of the Munsell colors," J. Opt. Soc. Am. **33**, 385 (1943).

TABLE IV. Absolute chroma limens for Munsell R, Y, G, B, and P, together with corresponding value limens, at value levels 2/ to 8/ as expressed in terms of Δx, Δy, ΔY.

Munsell	Origin			Obs. B			Obs. M			Obs. N			Mean		
	x	y	Y	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY	Δx	Δy	ΔY
Red															
N 2/	.3091	.3112	.0300	.0019	.0002	.0002	.0027	.0004	.0003	.0015	.0001	.0002	.0020	.0002	.0002
N 3/	.3082	.3121	.0651	.0002	.0000	.0000	.0010	.0000	.0001	.0004	.0003	.0001	.0005	.0001	.0001
N 4/	.3080	.3132	.1209	.0011	.0003	.0002	.0024	.0003	.0003	.0018	.0002	.0002	.0018	.0003	.0002
N 5/	.3084	.3140	.1906	.0020	.0003	.0002	.0017	.0002	.0002	.0027	.0002	.0003	.0021	.0002	.0002
N 6/	.3101	.3164	.3032	.0008	.0001	.0005	.0018	.0002	.0011	.0010	.0001	.0006	.0012	.0001	.0007
N 7/	.3086	.3154	.4433	.0018	.0010	.0003	.0015	.0001	.0005	.0008	.0001	.0003	.0014	.0004	.0004
N 8/	.3136	.3226	.5751	.0009	.0004	.0017	.0006	.0002	.0012	.0007	.0003	.0012	.0007	.0003	.0014
	Mean			.0012	.0003	.0004	.0017	.0002	.0005	.0013	.0002	.0004	.0014	.0002	.0005
Yellow															
N 2/	.3091	.3112	.0300	.0016	.0016	.0004	.0012	.0012	.0003	.0020	.0019	.0005	.0016	.0016	.0004
N 3/	.3082	.3121	.0651	.0013	.0012	.0001	.0014	.0018	.0002	.0014	.0018	.0002	.0014	.0016	.0002
N 4/	.3080	.3132	.1209	.0012	.0011	.0005	.0019	.0021	.0009	.0017	.0019	.0008	.0016	.0017	.0007
N 5/	.3084	.3140	.1906	.0014	.0015	.0008	.0017	.0019	.0010	.0022	.0025	.0014	.0018	.0020	.0011
N 6/	.3101	.3164	.3032	.0005	.0005	.0003	.0024	.0027	.0015	.0012	.0013	.0007	.0014	.0015	.0008
N 7/	.3086	.3154	.4433	.0011	.0013	.0001	.0020	.0023	.0002	.0007	.0008	.0002	.0013	.0015	.0002
N 8/	.3136	.3226	.5751	.0016	.0015	.0015	.0011	.0011	.0010	.0018	.0017	.0017	.0015	.0014	.0014
	Mean			.0012	.0012	.0005	0017	.0019	.0007	.0016	.0017	.0008	.0015	.0016	.0007
Green															
N 2/	.3091	.3112	.0300	.0011	.0015	.0003	.0009	.0011	.0002	.0006	.0004	.0001	.0009	.0010	.0002
N 3/	.3082	.3121	.0651	.0005	.0009	.0001	.0010	.0018	.0002	.0008	.0015	.0002	.0008	.0014	.0002
N 4/	.3080	.3132	.1209	.0009	.0017	.0004	.0009	.0017	.0004	.0005	.0013	.0003	.0008	.0016	.0004
N 5/	.3084	.3140	.1906	.0004	.0006	.0000	.0008	.0014	.0001	.0007	.0019	.0001	.0006	.0013	.0001
N 6/	.3101	.3164	.3032	.0006	.0010	.0000	.0009	.0014	.0000	.0005	.0009	.0000	.0007	.0011	.0000
N 7/	.3086	.3154	.4433	.0004	.0006	.0006	.0008	.0014	.0015	.0003	.0005	.0005	.0005	.0008	.0009
N 8/	.3136	.3226	.5751	.0007	.0010	.0004	.0006	.0008	.0003	.0007	.0010	.0004	.0007	.0009	.0004
	Mean			.0007	.0010	.0003	.0008	.0014	.0004	.0006	.0011	.0002	.0007	.0012	.0003
Blue															
N 2/	.3091	.3112	.0300	.0015	.0005	.0002	.0015	.0005	.0002	.0023	.0005	.0003	.0018	.0005	.0002
N 3/	.3082	.3121	.0651	.0008	.0004	.0001	.0030	.0017	.0003	.0008	.0004	.0001	.0015	.0008	.0002
N 4/	.3080	.3132	.1209	.0009	.0004	.0003	.0032	.0012	.0010	.0019	.0007	.0006	.0020	.0008	.0006
N 5/	.3084	.3140	.1906	.0019	.0009	.0000	.0023	.0010	.0001	.0027	.0013	.0001	.0023	.0011	.0001
N 6/	.3101	.3164	.3032	.0007	.0004	.0004	.0025	.0013	.0013	.0016	.0008	.0008	.0016	.0008	.0008
N 7/	.3086	.3154	.4433	.0006	.0001	.0007	.0018	.0004	.0021	.0007	.0001	.0009	.0010	.0002	.0012
N 8/	.3136	.3226	.5751	.0008	.0004	.0013	.0015	.0007	.0023	.0015	.0007	.0023	.0013	.0006	.0020
	Mean			.0010	.0004	.0004	.0023	.0010	.0010	.0016	.0006	.0007	.0016	.0007	.0007
Purple															
N 2/	.3091	.3112	.0300	.0012	.0043	.0002	.0011	.0042	.0003	.0001	.0065	.0001	.0008	.0050	.0002
N 3/	.3082	.3121	.0651	.0001	.0011	.0002	.0005	.0030	.0005	.0006	.0012	.0004	.0004	.0018	.0004
N 4/	.3080	.3132	.1209	.0003	.0022	.0002	.0005	.0027	.0003	.0003	.0011	.0001	.0004	.0020	.0002
N 5/	.3084	.3140	.1906	.0005	.0021	.0004	.0004	.0019	.0003	.0003	.0015	.0013	.0004	.0018	.0007
N 6/	.3101	.3164	.3032				.0007	.0025	.0015	.0005	.0015	.0011	.0006	.0020	.0013
N 7/	.3086	.3154	.4433	.0003	.0013	.0019	.0006	.0023	.0034	.0005	.0020	.0029	.0005	.0019	.0027
N 8/	.3136	.3226	.5751	.0006	.0018	.0011	.0005	.0014	.0008	.0003	.0009	.0005	.0005	.0014	.0008
	Mean			.0005	.0021	.0007	.0006	.0026	.0010	.0004	.0021	.0009	.0005	.0023	.0009
Value															
N 2/	.3091	.3112	.0300	.0000	.0000	.0002	.0004	.0002	.0004	.0000	.0000	.0003	.0000	.0000	.0003
N 3/	.3082	.3121	.0651	.0000	.0000	.0007	.0000	.0000	.0006	.0000	.0000	.0004	.0000	.0000	.0006
N 4/	.3080	.3132	.1209	.0000	.0000	.0010	.0000	.0000	.0019	.0000	.0000	.0010	.0000	.0000	.0013
N 5/	.3084	.3140	.1906	.0000	.0000	.0012	.0000	.0000	.0017	.0000	.0000	.0015	.0000	.0000	.0014
N 6/	.3101	.3164	.3032	.0000	.0000	.0012	.0000	.0000	.0038	.0000	.0000	.0024	.0000	.0000	.0025
N 7/	.3086	.3154	.4433	.0000	.0001	.0027	.0001	.0002	.0050	.0006	.0006	.0020	.0002	.0003	.0032
N 8/	.3136	.3226	.5751	.0000	.0002	.0042	.0000	.0002	.0044	.0000	.0002	.0050	.0000	.0002	.0044
	Mean			.0000	.0000	.0016	.0001	.0001	.0025	.0001	.0001	.0018	.0000	.0001	.0020

approximation of the directions in which the limens were taken to the true attributive directions, as estimated by the subcommittee. The approximation is usually rather good, as evident from the approach to parallelism between the liminal lines of hue and chroma and the contour lines of hue and chroma, respectively. Complete agreement would require the use, in the liminal observations, of color samples which conform exactly to the directions of the contours. Other characteristics of these diagrams may be mentioned. First, there is little relationship between magnitudes of limens and local spacings of contours. Second, the difference limens exhibit a strong tendency to increase in size with increase in distance from the achro-

matic point. Third, the absolute chroma limens, which are plotted as radiating out from the achromatic point in each chromaticity diagram, serve to suggest the minimum size of the visually neutral area under favorable conditions of viewing. This type of determination is similar to that of the least perceptible colorimetric purity.[6]

Since the I.C.I. system is essentially visual, at least with respect to the standard observer, indications of visual magnitude represented by I.C.I. specifications help to make the latter meaningful. For that reason the order of magnitude of liminal differences and variations is important. The liminal lines (divided by 5) in Fig. 1 give some indication of order of magnitude. More specifically, the average lengths of the liminal lines, computed by the distance formula from all the data of Tables I and II, are: hue 0.0023 and chroma 0.0035. The average Fechner fraction, or $\Delta Y/Y$ ratio was computed from the value data of Table III and proved to be 0.0087. The average ranges of variation (for the three observers) about these average liminal figures are: hue 0.0004–0.0079, chroma 0.0009–0.00129, and value 0.0019–0.0170. Of course the variations of the individual observers cover still greater ranges.

The actual differences in the trichromatic coefficients and Y figures which are entered in the tables seem remarkably small. In fact, a frequency tabulation showed that 40 to 50 percent of these delta-figures are less than 0.0010 in size, that is, are in the fourth or higher decimal place; and that virtually 100 percent do not encroach beyond the third decimal place.

Among the various sources of variation and error which have affected the reported figures, the most important are believed to be the observers themselves, the experimental conditions, and the spectrophotometric measurements. Some measure of experimental error would naturally be desirable and a place to look for such a measure is in figures which themselves should become vanishingly small under ideal experimental conditions. As mentioned above, the ΔY figures in Tables I and II seem largely ascribable to error, and the same is true of the ΔY figures in the first five sections of Table IV. Thus, while each of the hue or chroma limens in these tables was supposed to be taken at some fixed Munsell value, differences in the reflectance figures involved in the individual liminal determinations could largely account for the ΔY figures obtained. In this connection it may be noted that a comparison of reflectance data suggests a stimulational basis for a complication which was encountered during the original observations.[1] This complication consisted in perceptual variation of two of the attributes while the limen of the third attribute was being determined. Thus, the reflectance differences accompanying the chroma limens at value 2/ in Table II actually average somewhat larger than the corresponding reflectance figures for value limens in Table III. This could mean that sometimes it was easier to see value differences than chroma differences when determining chroma limens; and this could help account for the complication in question. The Δx and Δy figures in the value limens of Table III depend upon the real non-uniform relation between hue-chroma and chromaticity as well as on error; and so these Δx and Δy figures are part of the liminal specification.

In general it may be said that though efforts were made to minimize variability and error, the effects were judged to be sufficient to render useless the report of computations beyond the fourth place, even though many of the liminal figures appear to belong in that place. Perhaps this type of situation is to be expected as long as the normal observer remains the best available detector of small color differences.

[6] I. G. Priest and F. G. Brickwedde, "Minimum perceptible colorimetric purity as a function of dominant wavelength," J. Research Nat. Bur. Stand. **20**, 673 (1938).

Paraxial Differential Transfer Coefficients of a Lens System*

MARC G. DREYFUS**
Bausch & Lomb Optical Company, Rochester, New York
(Received March 26, 1948)

A shorter method is given for computing $C(u_i')$ and $C(y_i)$ for paraxial rays than that described by F. D. Cruickshank (see reference 1), and a calculation plan is given for computing the rate of change of position of an axial ray in the image space per change in curvatures, indices of refraction, and separations of lens elements.

IN lens production, quantitative knowledge of the rates of change of various aberrations of a lens system for changes in its variables is of considerable use, both in designing and in the determination of manufacturing tolerances. In this article, the curvatures of, and distances between the various spherical lens surfaces, and the indices of refraction of the lens elements are treated as variables; formulas are derived which can be applied to a ray trace to obtain numerical values for the change of various paraxial aberrations in the image per unit change in any one of the variables of the system. The formulas will be specifically applied in a plan for computing $\partial l_k'/\partial c_i$, $\partial l_k'/\partial_i d_{i-1}$, and $\partial l_k'/\partial N_i'$.

Since these partial derivatives represent tangents to curves they give, in general, valid approximations to the curves only in a small region about the point of tangency. As in F. D. Cruickshank's articles,[1,2] the traced ray furnishes the reference and the point of tangency. Therefore the paraxial differential transfer coefficients will give their most accurate results when the changes produced in the path of the ray are small, in which case the ray-path changes computed for small changes in several variables are approximately additive. Sign conventions used are those of A. E. Conrady in *Applied Optics and Optical Design*.[3]

Definitions: (See Fig. 1)

l_i = Traced ray's axial crossing distance from vertex of ith lens surface, before refraction at ith surface.
N_i = Index of refraction of material on object side of ith lens surface.
u_i = Traced ray's axial angle before refraction at ith lens surface.
y_i = Distance from axis to incidence point of traced ray on ith lens surface.
r_i = Radius of curvature of ith lens surface.
${}_id_{i-1}$ = Axial distance from ith to $(i-1)$th lens surface.
i_i = Incidence angle at ith lens surface.
k = Lens surface nearest image.
$n_i = N_i/N_i'$.

TABLE I. Axial differential transfer coefficient calculation plan.

	Regular plan for ith surface	Special formulas
(1) $= l_i'(2^-)$		$(1_k) = l_k'$
(2) $= (1)/l_i$	$= \operatorname{Lim}_{u_i \to 0}(u_i/u_k')$ Cf.*	
(3) $= n_i$	$= N_i/N_i'$	
(4) $= l_i - r_i$		
(5) $= [(3)-1](1)^2 N_i'$	$= -[1-n_i]l_i'^2(u_{i+1}/u_k')^2 N_i'$	
	$= -[1-n_i]l_i'(u_{i+1})/(u_k'^2) l_i u_i N_i' = \partial l_k'/\partial c_i$	
(6) $= (2)^2 N_i$	$= (u_i/u_k')^2 N_i = \partial l_k'/\partial_i d_{i-1}$	
(7) $= (1)(2)(4)/r_i$	$= (u_i/u_k')^2 l_i[(l_i - r_i)/r_i]$	(7P) $= -(1)(2)$
(8) $= (7^-) - (3)(7)$	$= \partial l_k'/\partial N_i'$	$(8_k) = -(3)(7)$

* A. C. Hardy and F. E. Perrin, *Principles of Optics* (McGraw-Hill Book Company, Inc., 1932), p. 43, Eq. 34.

* Research done at Bausch & Lomb Optical Company, Rochester, New York, under the supervision of Dr. A. A. Levin.

** Present address: Physics Department, Harvard University, Cambridge, Massachusetts.

[1] F. D. Cruickshank, J. Opt. Soc. Am. **36**: (a) p. 103, Eq. (3); (b) p. 16, Eq. (24); (c) p. 17, Eqs. (35), (38); (d) p. 17, Eqs. (37), (41); (e) p. 17, Eq. (40); (f) p. 17, Eqs. (25), (39) (1946).

[2] F. D. Cruickshank, Proc. Phys. Soc. London, **57**, pp, 302–310, 350–367, 419–435.

[3] A. E. Conrady, *Applied Optics and Optical Design* (Oxford University Press, London, 1929) (a) p. 45, line 4; (b) p. 7, Eq. (1).

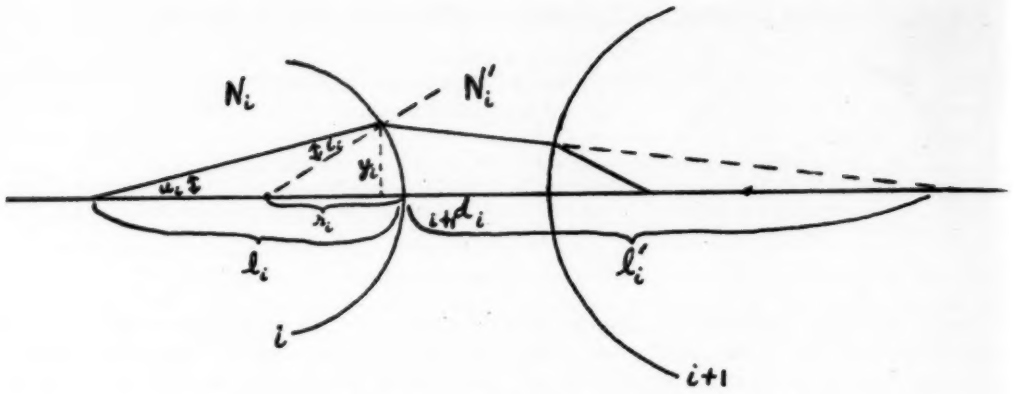


FIG. 1. Diagram of defined quantities.

Primed (′) quantities refer to values after crossing the lens surface going from object to image. For example:

l_i' = Traced ray's axial crossing distance from vertex of ith lens surface after refraction at ith surface.

N_i' = Index of refraction of material on image side of ith lens surface.

$$\Delta l_i = \Delta_i d_{i-1}$$

$$\therefore \quad \frac{\partial l_k'}{\partial_i d_{i-1}} = \frac{\partial l_k'}{\partial l_i} = \frac{N_i u_i^2}{N_k' u_k'^2} \qquad 3(a)$$

$$\frac{\partial l_k'}{\partial c_i} = \frac{\partial l_k'}{\partial l_i'} \frac{\partial l_i'}{\partial u_i'} \frac{\partial u_i'}{\partial c'}$$

$$= \frac{N_i' u_i'^2}{N_k' u_k'^2} \cdot \frac{-l_i'}{u_i} (1-n_i) l_i u_i \qquad 1(a), 1(b), 3(a)$$

F. D. Cruickshank derives the formulas:

$$\frac{\partial l_k'}{\partial c_i} = \frac{C(u_i')(\partial u_i'/\partial c_i)}{u_k'} \quad \text{where}$$

$$C(u_i') = \left(\frac{\partial y_k}{\partial u_i'} - l_k' \frac{\partial u_k'}{\partial u_i'} \right) \qquad 1(c)$$

$$= \frac{C(u_i')}{u_k'} (1-n_i) l_i u_i \qquad 1(b)$$

$$\frac{\partial l_k'}{\partial_i d_{i-1}} = \frac{C(y_i)}{u_k'} \frac{\partial y_i}{\partial_i d_{i-1}} \quad \text{where}$$

$$C(y_i) = u_k' \frac{\partial l_k'}{\partial y_i} \qquad 1(d)$$

$$= \frac{C(y_i)}{u_k'} \frac{\partial y_i}{\partial l_i} = \frac{C(y_i)}{u_k'} u_i.$$

Comparing the above formulas for $\partial l_k'/\partial c_i$ and $\partial l_k'/\partial_i d_{i-1}$ shows that:

$$C(u_i') = \frac{-N_i'}{N_k'} \frac{u_{i+1}}{u_k'} l_i' = \frac{-N_i'}{N_k'} l_i \frac{u_i}{u_k'}$$

and:

$$C(y_i) = \frac{N_i}{N_k'} \frac{u_i}{u_k'}.$$

With the aid of the latter formulas for $C(u_i')$ and $C(y_i)$, the formulas in F. D. Cruickshank's

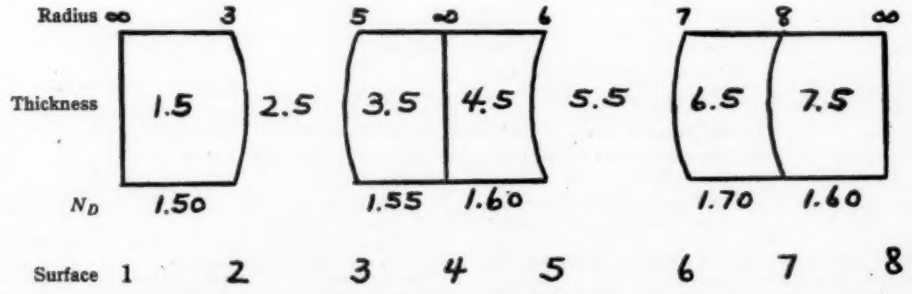


FIG. 2. Sample lens system.

TABLE II. Results of axial ray trace of sample lens system (Fig. 2).

Surface→	8	7	6	5	4	3	2	1
l_i	57.518	45.811	−14.814	−217.34	−206.19	−8.5000	−4.5000	−2.0000
l_k'	35.949	65.018	52.311	−9.3143	−212.84	−202.69	−6.0000	−3.0000

TABLE III. Axial differential transfer coefficient calculation for sample lens system (Fig. 2).

Surface→	8	7	6	5	4	3	2	1
1	35.949	40.636	46.402	29.174	28.570	28.085	19.825	13.217
2	.62500	.88703	−3.1322	−.13423	−.13856	−3.3042	−4.4056	−6.6083
3	1.6000	1.0625	.58824	1.6000	.96875	.64516	1.5000	.66667
4		37.811	−21.814	−223.34		−13.500	−1.5000	
5	775.38	165.13	−1507.2	510.69	−40.814	−433.84	196.52	−87.341
6	.62500	1.3376	9.8108	.02883	.02976	10.918	29.114	43.670
7	−22.468	170.37	452.93	145.77	3.9588	250.56	−43.670	87.341
8	35.949	−203.48	96.063	219.69	141.94	−157.69	316.06	−101.90

articles[1,2] which contain $C(u_i')$ and $C(y_i)$ can be put into a more easily computed form for paraxial rays. For example:

$$\frac{\partial l_k'}{\partial N_i'} = \frac{-\partial l_k'}{\partial n_i}\frac{N_i}{N_i'^2} + \frac{\partial l_k'}{\partial n_{i+1}}\frac{1}{N'_{i+1}} \qquad 1(e)$$

$$= \frac{C(u_i')}{u_k'}i_i\frac{N_i}{N_i'^2} - \frac{C(u'_{i+1})}{u_k'}i_{i+1}\frac{1}{N'_{i+1}} \qquad 1(f), 3(b)$$

$$= \frac{-N_i'}{N_k'}l_i\frac{u_i}{u_k'^2}\cdot\frac{(l_i-r_i)}{r_i}u_i\frac{N_i}{N_i'^2} + \frac{N'_{i+1}}{N_k'}l_{i+1}\frac{u_{i+1}}{u_k'^2}\frac{(l_{i+1}-r_{i+1})}{r_{i+1}}u_{i+1}\frac{1}{N'_{i+1}}$$

$$= \frac{1}{N_k'}\left[-n_i\left(\frac{u_i}{u_k'}\right)^2\frac{l_i-r_i}{r_i}l_i + \left(\frac{u_{i+1}}{u_k'}\right)^2\frac{l_{i+1}-r_{i+1}}{r_{i+1}}l_{i+1}\right].$$

In the following sample calculation plan for $\partial l_k'/\partial N_i'$, $\partial l_k'/\partial_i d_{i-1}$ and $\partial l_k'/\partial c_i$, bracketed, unminused numbers (2) refer to previously calculated lines for the lens surface being calculated. Bracketed, minused numbers (2⁻) refer to previously calculated lines for the neighboring surface on the image side of the surface being calculated. The calculation starts at the lens surface nearest the image, where special formulas (1_k), and (8_k) are used instead of the regular formulas (1), and (8). The calculation then proceeds surface by surface to the lens surface nearest the object. At a plano surface formula (7P) is used instead of the regular formula (7), and formula (4) may be omitted. It can be seen that an error in computing any line in the ith surface calculation except lines 1 and 2 will, at worst, only cause an error in the $(i-1)$ surface calculation. It is assumed that $N_k'=1$. In Table III line 5 is $(\partial l_k'/\partial c_1)$, line 6 is $\partial l_k'/\partial_i d_{i-1}$, and line 8 is $\partial l_k'/\partial N_i'$.

Irradiance Distribution in a Lloyd Mirror Interference Pattern*, **

ROBERT N. WOLFE AND FRED C. EISEN
Eastman Kodak Company, Rochester, New York

(Received March 18, 1948)

The Lloyd mirror interference pattern reappears and disappears in a periodic manner at distances from the mirror edge greater than the distance to the initial point of disappearance. In certain cases bright lines appear where, according to the textbook theory, which assumes spectrally homogeneous radiation and a slit source of infinitesimal width, dark lines should appear. A photograph of such a pattern is shown, measured values of the relative irradiance as a function of distance from the mirror edge are presented, and a theoretical determination which accounts for certain features of this variation of irradiance with distance is outlined.

If we let I_R = relative irradiance at a distance P from edge of mirror, θ = angular tilt of mirror, α = angular subtense of slit, ν_0 = wave number of radiation, ν_1 = width of wave band, and $\omega = 2\pi\nu_0$, then

$$\begin{aligned} I_R = 2\omega\alpha(\nu_1/\nu_0) &- (1/P)\mathrm{Si}[P\omega(1+\tfrac{1}{2}(\nu_1/\nu_0)(2\theta+\alpha)] \\ &+ (1/P)\mathrm{Si}[P\omega(1-\tfrac{1}{2}(\nu_1/\nu_0)(2\theta+\alpha)] \\ &+ (1/P)\mathrm{Si}[P\omega(1+\tfrac{1}{2}(\nu_1/\nu_0)(2\theta-\alpha)] \\ &- (1/P)\mathrm{Si}[P\omega(1-\tfrac{1}{2}(\nu_1/\nu_0)(2\theta-\alpha)]. \end{aligned}$$

If spectrally homogeneous radiation is assumed, then $I_{R1} = 0.50 - 0.50\cos 4\pi P\theta\nu_0(\sin 2\pi P\nu_0\alpha/2\pi P\nu_0\alpha)$.

INTRODUCTION

THE Lloyd mirror system for producing optical interference is well known, and discussions of it appear in standard textbooks[1] on physical optics. However, the reappearance of the interference lines as the order of interference is increased, after their disappearance, as a result of the lack of spectral homogeneity of the radiation and finite size of source, does not seem to be so well known. Also, a search of the literature has not disclosed an experimental measurement or a theoretical determination of the spatial irradiance distribution in this type of interference pattern. The purpose of this paper is to report both experimental measurement and theoretical determination of the spatial irradiance distribution in such an interference pattern.

EXPERIMENTAL

Figure 1 illustrates the essential features of the Lloyd mirror. It consists of a plane front surface mirror (m) which forms a virtual image of the radiation source, S_A, at S_B. Optical interference takes place between the radiation from the source, S_A, and its image, S_B. A photographic plate (p) placed at the end of the mirror (m) will record the interference taking place at I in the plane of the plate. The point I is at a distance, P, from the rear edge of the mirror. The distance between adjacent bands of the interference pattern is determined by the angle (θ) at which the mirror is tilted relative to a line joining the point of incidence, M, and the source. Lines S_AI and S_AM diverge from S_A at an angle ϕ which, in practice, is small.

Figure 2 shows a photographic record of the interference pattern obtained with this system. The radiation source was a slit 22 microns in width irradiated with a high pressure type H-6 mercury arc which was cooled by means of air jets. The radiation was filtered to obtain a spectral band in the neighborhood of wave-length 5461A. Spectrally homogeneous radiation could not be obtained from the H-6 arc because the spectrum lines are broadened as a result of the high pressure. The photographic record was made on a type 548GH spectroscopic plate placed 157 mm from the slit and exposed for 3 minutes. It was developed in a solution consisting of two parts Kodak D-8 developer stock solution and one part water for 1.5 minutes at 68°F.

At the left of the interference pattern in Fig. 2 is a dark area which was shielded from exposure by the end of the mirror. A dark line, which cannot be seen in the figure, is formed at the edge of the mirror because of the phase change of π radians introduced in the reflected beam on reflection from the mirror surface. As

* Communication No. 1184 from the Kodak Research Laboratories.

** Presented at the New York Meeting of the Optical Society of America, March 4–6, 1948.

[1] F. A. Jenkins and H. E. White, *Fundamentals of Physical Optics* (McGraw-Hill Book Company, Inc., New York, 1937), p. 66.

the distance to the right of the edge of the mirror is increased, the ratio of the irradiance of the light lines to that of the dark lines decreases to approximately unity, increases, and then decreases again. Two or possibly three positions can be noted where the ratio is unity, and between these positions groups of lines can be seen. Where the ratio is unity, it can be said that the lines disappear. Also, as will be discussed later, if the distance between the lines is S, then the distance from the edge of the mirror to any dark line is as follows:

$$\begin{aligned} &\text{1st group}—n S, \\ &\text{2nd group}—(n+\tfrac{1}{2})S, \\ &\text{3rd group}—n S, \text{ etc.}, \end{aligned}$$

where n is the ordinal number of the line, counting from the edge of the mirror.

The spatial irradiance distribution in an interference pattern having a greater distance between the lines than the pattern in Fig. 2 was measured. This was done by imaging the pattern formed in the plane normally occupied by the photographic plate onto a neutral colloidal carbon wedge[2] with a well-corrected lens, following the procedure outlined by Jones and Wolfe.[3] The wedge was oriented so that its isodensity lines were perpendicular to the lines of the interference pattern. The photographic material was placed immediately behind and in contact with the wedge. Figure 3 shows the resulting photographic image. A plot of log irradiance against distance from the mirror edge was then obtained by making an isodensity record of the resulting photographic image. The features of the interference pattern in Fig. 2 can be seen in this photographic image (Fig. 3).

FIG. 2. Lloyd mirror interference pattern.

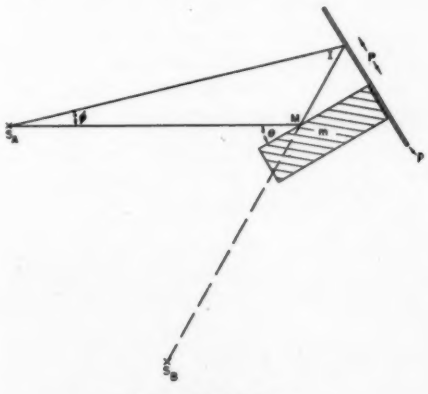


FIG. 1. Lloyd mirror.

[2] C. E. K. Mees and S. H. Wratten, "The wedge spectrograph," Brit. J. Phot. **54**, 384 (1907).

[3] L. A. Jones and R. N. Wolfe, "A method for the measurement of the energy distribution in optical images," J. Opt. Soc. Am. **35**, 559 (1945).

THEORETICAL

A mathematical expression accounting for the features of the Lloyd mirror interference pattern just discussed will now be derived.

In order to simplify the mathematics it will be assumed that:

(a) The mirror reflects all radiation falling on it at all angles of incidence. Thus, the source, S_B, can be considered as being equal in intensity to the source, S_A.

(b) The mirror has no effect on the polarization of the radiation striking it, i.e., the elliptical polarization introduced by the mirror is neglected.

(c) There is no extraneous or scattered light in the system.

(d) The angular separation of source and image is small.

(e) The distance from the axis of the system to points under consideration is small.

(f) The slit plane is maintained perpendicular to the plane of the mirror surface.

Also, the mathematics will be simplified further by considering only the two-dimensional aspects of the problem.

A detailed diagram of the Lloyd mirror system is shown in Fig. 4. This is a sectional view of the system taken in a plane perpendicular to the rear edge of the mirror, the edge being normal to the plane of the paper at O. The line OO' in the plane of the mirror is considered as the axis of the system. The line CD, which is in the slit plane, is perpendicular to OO'. The line Op is also perpendicular to OO' and is at a distance, b, from CD. The slit has a width, W, and a length, l. The slit length is small compared with the distance, b. The angular tilt of the mirror is θ, and

α is the angular subtense of the slit. The discrepancy between θ in Fig. 1 and θ in Fig. 4 is negligible because of assumptions (d) and (e).

Consider as a source of radiation an element of the slit at A', a distance, Z, from the axis having a width, dZ, and a length, l. This element has a mirror image at B'. Let the wave number of the radiation at the center of the frequency band used be ν_0, and let the band width of the radiation be ν_1. The problem is to determine the spatial irradiance distribution along line Op due to the integrated effect of the radiation of band width, ν_1, coming from the slit of width W.

Now, consider point I' on line Op and at a distance, P, from O. The resultant irradiance at I' depends on the phase relations existing between the radiation arriving from A' and that arriving from B'. This, in turn, depends on the geometrical path difference between $A'I'$ and $B'I'$. Let this path difference be designated by Δ'.

Now,

$$\Delta' = B'I' - A'I' = [(Z+P)^2+b^2]^{\frac{1}{2}} - [(Z-P)^2+b^2]^{\frac{1}{2}}.$$

Transforming this, it becomes

$$\Delta' = \frac{4ZP}{((Z+P)^2+b^2)^{\frac{1}{2}} + ((Z-P)^2+b^2)^{\frac{1}{2}}}.$$

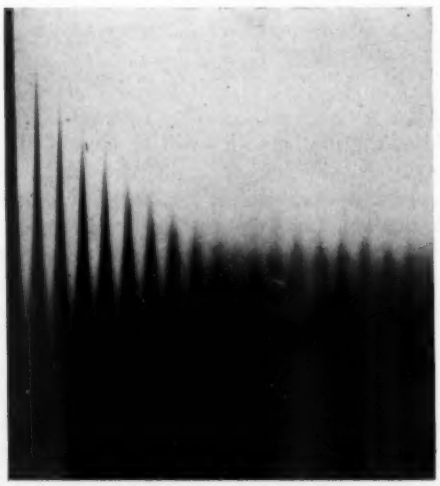


FIG. 3. Lloyd mirror wedge interference pattern showing variation of log irradiance with distance.

If both $(Z+P)$ and $(Z-P)$ are less than 10 percent of b, i.e., applying assumption (d), then

$$\Delta' = 2ZP/b \tag{1}$$

to within 1 percent.

At I', waves of radiation from A' and B' are falling, and these are out of phase by an amount determined by the distance, Δ'. The resultant amplitude of the combination of the two waves[4] is a product of a cosine function of Δ' and a cosine function of time.

The maximum amplitude ($\eta_{\max}$) that can occur at any point on line Op is its value when the cosine factor which is a function of time equals unity. This is

$$\eta_{\max} = 2r\cos\pi\Delta'\nu.$$

The irradiance at any point is proportional to the square of this amplitude. Thus,

$$I = 4Kr^2\cos^2\pi\Delta'\nu, \tag{2}$$

where I is the irradiance in watts per square centimeter on the exposure plane per square centimeter of source area. Taking into account the phase change of π at reflection, and making use of Eq. (1), we see that

$$I = 4Kr^2\sin^2 2\pi(P/b)\nu Z. \tag{3}$$

The resulting irradiance is the integral of all the elements of irradiance emitted from the various elemental frequency bands and elemental areas of the slit. The elemental area is $l\,dZ$, in which l is constant and is short enough so that Δ' can be considered constant for any point in the area.

If we let I_1 now be the integral of all these irradiance values, we can write

$$I_1 = \int_{\nu_0-(\nu_1/2)}^{\nu_0+(\nu_1/2)} \int_{Z_1-(W/2)}^{Z_1+(W/2)} Il\,dZ\,d\nu,$$

where the distance from the axis to the center of the slit is Z_1. Substituting for I from Eq. (3) and dividing by the constants, this becomes

$$\frac{I_1}{4Kr^2l} = \int_{\nu_0-(\nu_1/2)}^{\nu_0+(\nu_1/2)} \int_{Z_1-(W/2)}^{Z_1+(W/2)} \sin^2 2\pi(P/b)\nu Z\,dZ\,d\nu. \tag{4}$$

[4] R. A. Houstoun, *A Treatise on Light* (Longmans, Green and Company, London, 1915), p. 132.

Integrating with respect to Z, this becomes

$$\frac{I_1}{4Kr^2l}=\int_{\nu_0-(\nu_1/2)}^{\nu_0+(\nu_1/2)}[(1/4\gamma)(2\gamma Z-\sin 2\gamma Z)]_{Z_1-(W/2)}^{Z_1+(W/2)}d\nu,$$

where $\gamma=2\pi(P/b)\nu$. Substituting the limits for Z and using the trigonometric relations for the sine of the sum and the sine of the difference of two angles, we have

$$\frac{I_1}{2Kr^2lW}=\int_{\nu_0-(\nu_1/2)}^{\nu_0+(\nu_1/2)}1-\frac{\cos 2\gamma Z_1\sin\gamma W}{\gamma W}d\nu. \quad (5)$$

Substituting for γ as defined previously, we get

$$\frac{I_1}{2Kr^2lW}=$$

$$\int_{\nu_0-(\nu_1/2)}^{\nu_0+(\nu_1/2)}1-\frac{\cos(4\pi P\nu Z_1/b)\sin(2\pi P\nu W/b)}{(2\pi P\nu W/b)}d\nu.$$

Let

$$4\pi(P/b)Z_1=H,\quad 2\pi W/b=J,$$

and $$I_2=(I_1/2Kr^2lW),$$

then

$$I_2=\nu_1-\int_{\nu_0-(\nu_1/2)}^{\nu_0+(\nu_1/2)}\frac{\cos H\nu\sin JP\nu}{JP\nu}d\nu.$$

Now,

$$\cos x\sin y=(\sin(x+y)/2)-(\sin(x-y)/2).$$

Using this, we have

$$I_2=$$

$$\nu_1-\int_{\nu_0-(\nu_1/2)}^{\nu_0+(\nu_1/2)}\left[\frac{\sin(H+JP)\nu}{2JP\nu}-\frac{\sin(H-JP)\nu}{2JP\nu}\right]d\nu.$$

Upon separating the integrals, we have

$$I_RP=2JP\nu_1-\int_0^{[\nu_0+(\nu_1/2)](H+JP)}(\sin u'/u')du'$$

$$+\int_0^{[\nu_0-(\nu_1/2)](H+JP)}(\sin u'/u')du'$$

$$+\int_0^{[\nu_0+(\nu_1/2)](H-JP)}(\sin u''/u'')du''$$

$$-\int_0^{[\nu_0-(\nu_1/2)](H-JP)}(\sin u''/u'')du'', \quad (6)$$

where

$$u'=(H+JP)\nu,\quad du'=(H+JP)d\nu,$$

and

$$u''=(H-JP)\nu,\quad du''=(H-JP)d\nu,$$

and

$$I_R=2JI_2.$$

These integrals are of the form of the sine integral

$$\int_0^p\frac{\sin u}{u}du.$$

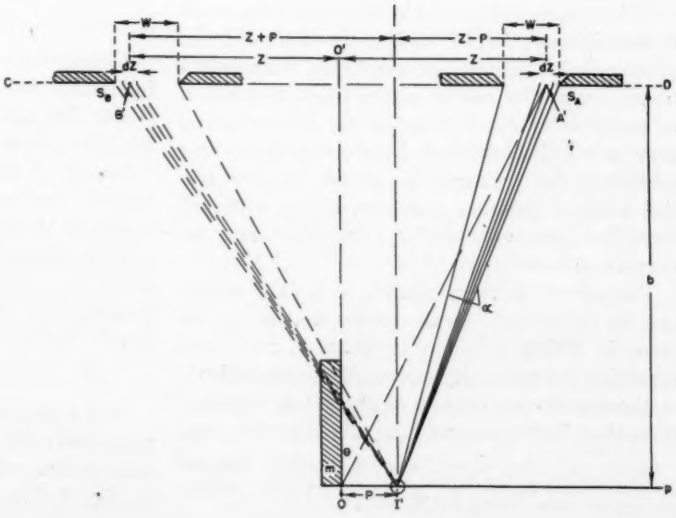


FIG. 4. Detailed diagram of Lloyd mirror interference system.

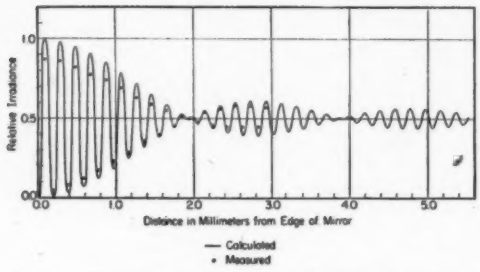


FIG. 5. Theoretical calculation and experimental measurement of irradiance distribution in a Lloyd mirror interference pattern.

Let

$$\int_0^{\rho}\frac{\sin u}{u}du=\mathrm{Si}(\rho) \quad \text{and}$$

$$\omega=2\pi\nu_0.$$

Also, on the basis of assumptions (d) and (e), we can write

$$Z/b=\theta \text{ in radians,}$$

and

$$W/b=\alpha \text{ in radians.}$$

On making these substitutions and the substitutions for H and J, as previously defined, Eq. (6) becomes

$$\begin{aligned}I_R=&2\omega\alpha(\nu_1/\nu_0)\\&-(1/P)\mathrm{Si}[P\omega[1+\tfrac{1}{2}(\nu_1/\nu_0)](2\theta+\alpha)]\\&+(1/P)\mathrm{Si}[P\omega[1-\tfrac{1}{2}(\nu_1/\nu_0)](2\theta+\alpha)]\\&+(1/P)\mathrm{Si}[P\omega[1+\tfrac{1}{2}(\nu_1/\nu_0)](2\theta-\alpha)]\\&-(1/P)\mathrm{Si}[P\omega[1-\tfrac{1}{2}(\nu_1/\nu_0)](2\theta-\alpha)]. \quad (7)\end{aligned}$$

This equation gives the intensity as a function of the distance, P, the angular tilt of the mirror, θ, the angular subtense of the slit, α, the spectral frequency at the center of the frequency band, ν_0, and the band width, ν_1. It can be evaluated over only a limited range by the use of existing tables[5] of sine integrals. From the form of the sine integral function, however, it can be noted that, for large values of P, the irradiance becomes constant.

The nature of the variation of I as P varies can be determined more readily in the special cases in which either ν_1 approaches zero (the condition for spectrally homogeneous radiation), or the angular subtense α of the slit approaches zero, than in the general case where both ν_1 and α are finite. In the case of spectrally homogeneous radiation and a finite slit width, Eq. (4) can be written

$$I_2\frac{W}{2}=\int_{Z_1-(W/2)}^{Z_1+(W/2)}\sin^2 2\pi(P/b)\nu_0 Z\,dZ. \quad (8)$$

Following the integration procedure used in obtaining Eq. (5), substituting for Z_1, b, and W, as discussed above, and substituting I_{R1} for $0.5I_2$, where I_{R1} is the relative irradiance adjusted so that the maximum possible irradiance would be unity, Eq. (8) becomes

$$I_{R1}=0.5-0.5\cos 4\pi P\theta\nu_0(\sin 2\pi P\nu_0\alpha/2\pi P\nu_0\alpha). \quad (9)$$

An analogous equation for the case of an infinitesimal slit width and a finite frequency band is

$$\begin{aligned}I_{R2}=&0.5-0.5\cos 4\pi P\theta\nu_0\\&\times(\sin 2\pi P\theta\nu_1/2\pi P\theta\nu_1), \quad (10)\end{aligned}$$

where

$$I_{R2}=I_2(W/2\nu_1).$$

A graph of I_{R1} as a function of P is shown in Fig. 5. This was calculated by means of Eq. (9), using the values of the experimental constants for θ, ν_0, and α that were used in obtaining Fig. 3. These constants are:

$$\begin{aligned}\theta&=4.8 \text{ minutes}\\&=1.4\times10^{-3} \text{ radian,}\\\nu_0&=1.83\times10^4 \text{ cm}^{-1},\\\alpha&=29 \text{ seconds}\\&=0.14\times10^{-3} \text{ radian.}\end{aligned}$$

Consider Eqs. (9) and (10). It can be seen by inspection that the expression in each case is a cosine function of θ, in which the right-hand member of the equation oscillates about a mean value of 0.5 and which is modulated by a function of the form $(\sin x/x)$. Thus, the irradiance varies in the manner indicated in Fig. 5 for the case of spectrally homogeneous radiation and a finite slit width, and also for the case of an infinitesimal slit width and a finite frequency band (Eq. 10).

DISCUSSION

The experimental points on Fig. 5 represent measurements of the irradiance at the maxima and minima of the interference pattern shown in Fig. 3. The positions of the measured points

[5] *Tables of Sine, Cosine and Exponential Integrals* (Federal Works Agency, WPA, City of New York, Project No. 765-97-3-10, 1940), Vol. I, Vol. II.

relative to the edge of the mirror agree closely with the positions of the calculated maxima and minima, and the measured irradiance values agree qualitatively with the calculated values.

If P_1 is the distance from the edge of the mirror to one dark band and P_2 is the distance to the next dark band, then

$$P_2-P_1=S,$$

where S is the distance between dark bands. A dark band occurs wherever

$$\cos 4\pi P\theta\nu_0=1$$

if the sine term is positive. Thus,

$$4\pi P_1\theta\nu_0=2n\pi$$

and

$$4\pi P_2\theta\nu_0=2(n+1)\pi.$$

Solving for S, we have

$$S=\frac{1}{2\theta\nu_0},$$

which is the well-known expression for the line spacing in a Lloyd mirror interference pattern. Calculating for the value of S in Fig. 5, we have

$$S=1/(2\times 1.40\times 10^{-3}\times 1.83\times 10^{4})\text{ cm}$$
$$=0.195\text{ mm}.$$

As can be seen, the irradiance of the bright lines decreases and that of the dark lines increases as the distance from the edge of the mirror increases. They both approach a value of 0.50, at which point they disappear.

The distance from the edge to the first point of disappearance can be calculated in the following way:

Let P_α be the distance from the edge of the mirror to the point where the lines disappear. This occurs when the sine term in Eq. (9) becomes zero. Thus,

$$\sin 2P_\alpha\pi\nu_0\alpha=0,$$

or

$$2P_\alpha\pi\nu_0\alpha=\pi.$$

Solving for P_α, we have

$$P_\alpha=\frac{1}{2\nu_0\alpha}. \tag{11}$$

Therefore,

$$P_\alpha=1/(2\times 1.83\times 10^{4}\times 0.14\times 10^{-3})\text{ cm}$$
$$=1.95\text{ mm}.$$

This, it can be seen, agrees closely with the measured value which means that, under the conditions of the experiment, the lack of spectral homogeneity in the radiation which was used does not influence the irradiance distribution.

It is of interest to derive the expression for the distance from the edge to the point of disappearance, using Eq. (10) for which an infinitesimal slit width and a finite frequency band are assumed. Let $P_{\nu 1}$ be this distance. Then, by analogy with the previous calculation, we can write

$$P_{\nu 1}=\frac{1}{2\theta\nu_1}. \tag{12}$$

Equation (12) indicates that $P_{\nu 1}$ is inversely proportional to the tilt angle of the mirror (θ). Thus, by increasing θ, $P_{\nu 1}$ can be decreased. Consequently, it would be expected that, for larger values of θ than were used in obtaining the interference pattern shown in Fig. 3, the finite frequency band would make the distance from the edge of the mirror to the point of disappearance of the lines less than 1.95 mm. Preliminary observations indicate that this is true.

Where the sine term is negative, as is the case for angles in the third and fourth quadrants, the positions of the light and dark lines are interchanged. This is the case for the group of lines between the first and second points of disappearance, as was mentioned previously. The sine term is positive for the next group of lines. Thus, the theory indicates that the lines will disappear and then reappear with reversal of "phase" as P increases.

The denominator under the sine term is directly proportional to P. Because of this, the relative irradiance values for successive groups of lines will depart less and less from 0.5 and will approach a constant value of 0.5 as P approaches infinity. This reduction in "amplitude of oscillation" as P increases can be seen in the theoretical curve (Fig. 5).

ACKNOWLEDGMENT

The authors wish to thank Dr. G. C. Higgins and Dr. M. Herzberger, of these Laboratories, for their assistance in preparing the theoretical section of this paper.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 8 AUGUST, 1948

Recommended Footcandle Levels for Prolonged Critical Seeing

MATTHEW LUCKIESH
Lighting Research Laboratory, General Electric Company, Nela Park, Cleveland
(Received May 5, 1948)

Seeing is a complex human activity involving more than the visual sense and its ability to convert the stimulus, radiant energy, into sensations of brightness and color. Seeing is accomplished by the human being operating as a human seeing-machine. Therefore, efficient, certain, comfortable, and easy seeing involves psychophysiological factors and effects of seeing as well as the psychophysical characteristics of the visual task and its environment. There is some confusion and some difference of opinion in regard to levels of illumination that are now being recommended. This is not surprising inasmuch as recommended footcandle levels are very generally compromises involving many aspects of economics, practicability, and conservation. Adequate specifications of light, lighting, brightness, and color do not arise from a single study or even a limited group of studies. Pertinent data are yielded by researches along various major avenues involving achievement, conservation of human resources, subnormal as well as normal vision, supra-threshold as well as threshold visibility, and proper considerations of footcandle level as a means to the ends which are brightness level and visibility level. The author presents data and analyses which reply specifically to certain criticisms and which, he believes, support the conclusion that footcandle levels recommended by the Illuminating Engineering Society, as well as by himself and colleagues, are conservative in relation to the indicated ideal levels for various tasks of prolonged critical seeing.

THE science of seeing, upon which a science of lighting must rest, is now a formidable structure of knowledge. It may be likened to a complex building sufficiently advanced in construction so that the major features are well defined and many minor details can be distinguished. As the author looks backward over nearly four decades of continuous research by himself and colleagues in the extensive and related fields of light, color, lighting, vision, and seeing, several major avenues of investigation stand out rather clearly. The contributions of light to the complex human activity of seeing have been studied by determining its relationships to (1) human achievement, (2) conservation of human and material resources, (3) the limitations and abilities of the visual sense, (4) subnormal as well as normal vision, (5) the visibility of various visual objects and complex visual tasks, (6) supra-threshold visibility levels in contradistinction to threshold visibility levels, (7) "factors of safety" which are as justifiable in seeing as in bridge design, and to (8) ease of seeing which involves any unnecessary or undesirable waste of human resourses as largely determined by psychophysiological effects of seeing. These descriptions of major avenues of interest are merely glimpses of the complexity and ramifications of light in its effect upon the efficacy, efficiency, safety, comfort, and welfare of human beings operating as *human seeing-machines*.

During nearly forty years of continuous research along these various avenues, it has become obvious to us that no single study or even a limited group of studies is all-important and conclusive. Nevertheless, newcomers in this complex field of seeing can easily make the error of oversimplifying a matter which is inherently complex, of ignoring or misinterpreting the results of researches in other avenues of seeing, and of extending their results obtained in a restricted area beyond justifiable limits.

In a recent paper, Simonson and Brozek[1] published the results of studies of the effects of level of illumination on visual performance and fatigue in which they make some of these errors. In their discussion of their results, they state: "We wish to emphasize that the specific conclusions arrived at in this study are limited to our type of visual task and subjects." Notwithstanding this statement near the end of their paper, they leave unmodified their statements in the early paragraphs and throughout their paper to the effect that footcandle levels recommended by official committees of the Illuminating Engineering Society are too high and unjustified.

[1] E. Simonson and J. Brozek, "Effects of illumination level on visual performance and fatigue," J. Opt. Soc. Am. **38**, 384 (1948).

VISUAL TASKS OF HIGH CONTRAST

Reading matter printed with so-called black ink on so-called white paper is the most common visual task in our present civilization. Such tasks are of inherently high visibility owing to the high brightness contrast between the critical details and their background. However, these tasks are not representative of uncounted other tasks throughout the work-world as, for example, sewing with black thread on dark textiles, which are of relatively low visibility. Investigations of the visual sense for many years were confined to black characters on white backgrounds. Even the test charts and other test objects of the eyesight specialist are generally of this specialized character. They serve their restricted purpose, but they have dominated too much the thinking of many investigators and practitioners.[2] The writer and his colleagues[3] recognized this fact many years ago and turned their attention to studies of the relationships of fundamental factors such as size, brightness contrast, brightness level, and time, which are involved in the performance of any visual task. However, for the present we shall confine the discussion to visual tasks whose critical details of relatively low reflectance are viewed against backgrounds of relatively high reflectance.

In the study reported by Simonson and Brozek, they used as their "work test" small letters printed (presumably with so-called black ink) on dull so-called white paper which was mounted on a belt driven at a uniform speed by an electric motor. Each letter was exposed for 0.56 second as it passed a narrow slit. The subjects, who possessed average normal vision, copied the letters by hand in their order of appearance while keeping their eyes focused on the slit in which the letters successively appeared. Obviously, this part of the study involved accuracy of recognition and falls in the category of achievement. They do not describe certain factors which affect the seeing conditions, and we must accept their data which cannot be separately appraised because of the absence of observed data from their paper. Their own appraisal of their results in regard to optimum footcandle levels is stated in their own words as follows:

> "Although for the average performance the optimum at 100 footcandles was not statistically significant, it fits so well into the general pattern of the other performance criteria that there is little doubt of the reality of a performance optimum above 50 footcandles but below 300 footcandles. Both performance-drop from start to end of the work period and performance-range show a definite optimum at 100 footcandles. For the blinking-rate average there was a well established optimum at 100 footcandles."

These results are in substantial agreement with results published by Luckiesh and Moss whose work they criticize. However, of more importance is the fact that their own data disprove their contention that the footcandle levels recommended by the I.E.S. are too high.

In the latest I.E.S. *Recommended Practice of Office Lighting*,[4] 50 footcandles are recommended for "difficult seeing" or tasks involving small type or critical details, poor contrast, and prolonged periods of time such as are encountered in auditing, accounting, drafting, etc. In the latest I.E.S. *Recommended Practice of Home Lighting*,[5] 40 footcandles are recommended for prolonged reading of small type. In the I.E.S. *Recommended Practice of Industrial Lighting*, 100 footcandles or more are recommended for some specific tasks involving (a) discrimination of extremely fine details under conditions of (b) extremely poor contrast and (c) for long periods of time. It is difficult to see why Simonson and Brozek have criticized the I.E.S. recommendations, inasmuch as they concluded from their own data that the optimum level of illumination was between 50 and 300 footcandles with the weight of evidence at about 100 footcandles. Apparently their work test consisted of small black letters on a white background but involved neither subnormal vision nor prolonged critical seeing comparable to many hours daily. Their results do not support their criticisms of the I.E.S. recommendations and the results of comparable researches by Luckiesh and Moss.

---

[2] Matthew Luckiesh, "Visual acuity and visual tasks," Illum. Eng. **39**, 415 (1944); "Test-charts representing a variety of visual tasks," Am. J. Ophthal. **27**, 270 (1944).

[3] P. W. Cobb and F. K. Moss, "The four variables of the visual threshold," J. Frank. Inst. **205**, 831 (1928); "Four fundamental factors in vision," Trans. Illum. Eng. Soc. **23**, 496 (1928).

[4] *Recommended Practice of Office Lighting* (Illuminating Engineering Society, New York, 1947), p. 8.

[5] *Recommended Practice of Home Lighting* (Illuminating Engineering Society, New York, 1945) p. 8.

BRIGHTNESS LEVEL

Actually, footcandle level by itself has little meaning in seeing. Footcandle level is a factor in producing brightness level, but the latter depends just as much upon the diffuse reflectance of the incident light. Brightness level is directly important in seeing,[6–8] and visibility level is still more revealing. However, for the present let us examine brightness level as a product of footcandle level and diffuse reflectance. If the letters of the work test used by Simonson and Brozek had been perfectly black and printed on a dark gray background whose diffuse reflectance was one-tenth that of the dull white paper which they actually used, they would have found the optimum level of illumination to be between 500 and 3000 footcandles. In other words, it would require this range of footcandle levels to produce the same brightness levels on the dark gray background as those of the white paper when the levels of illumination ranged from 50 to 300 footcandles. There is no possible argument against this simple fact of diffuse brightness level being the result of a combination of footcandle level and diffuse reflectance. If there were no extraneous factors in the surroundings or seeing conditions which seriously affected the results obtained by Simonson and Brozek, they would have obtained an optimum in the range of 500 to 3000 footcandles if the letters had been printed perfectly black on a dark gray paper having a diffuse reflectance of one-tenth that of the white paper they used. Actually, their letters were probably not black but nearly black, for so-called black ink commonly has a diffuse reflectance of at least 4 percent. This lower contrast between the near-black letters and the dark gray paper would require more than 500 to 3000 footcandles to raise the letters to the same visibility levels as they were when printed on the white paper and viewed under 50 to 300 footcandles.

This is a simple straightforward example of the importance of brightness level and of the influence of low brightness contrasts on visibility. In the I.E.S. *Recommended Practice of Home Lighting*,[5] "100 footcandles or more" are recommended for sewing on dark goods involving fine needlework. Similar recommendations are made for other tasks, involving fine details of low contrasts with their background, in the I.E.S. *Recommended Practice of Industrial Lighting*.[9] Certainly these brief analyses reveal that the I.E.S. recommendations of footcandle levels are not only conservative, but for many of the tasks of inherently low visibility they are ultra-conservative. Lack of comprehension of the importance of brightness level as compared with footcandle level has not only led astray newcomers in this complex field of seeing, but it has deterred the recommendation and application of much higher footcandle levels in many places where prolonged critical seeing is done for hours daily with tasks of inherently low visibility. The inherently low visibility of objects and tasks whose critical details are of low brightness-contrast with their backgrounds has not yet received the attention it deserves throughout the realms of critical seeing. A major aim of the science of lighting is the specification of footcandles to fit the task. Certainly this is an objective as worthy as the well established practice of fitting motors to mechanical tasks. Human seeing-machines have been long neglected in this respect and one of the various objectives of the author and his colleagues has been to develop a practical means of accomplishing this. Many years of research underlie the development of the Luckiesh-Moss visibility meter and its calibration and use.[10–12] At the present time this appears to be the only practicable method of determining supra-threshold visibilities of visual objects and tasks and of determining footcandle levels necessary to raise them to a given or desirable supra-threshold visibility level. This involves more than visual acuity, for this method

[6] Matthew Luckiesh, *Light, Vision and Seeing* (D. Van Nostrand Company, Inc., New York, 1944), pp. 68–86, 106–124.

[7] Matthew Luckiesh and A. A. Eastman, "Footcandles for critical seeing," Illum. Eng. **41**, 828 (1946).

[8] Matthew Luckiesh, "Footcandle levels—threshold, ideal, optimum and recommended," Illum. Eng. **43**, 395 (1948).

[9] *Recommended Practice of Industrial Lighting* (Illuminating Engineering Society, New York, 1942), p. 29.

[10] Matthew Luckiesh and F. K. Moss, "Visibility—its measurement and significance in seeing," J. Frank. Inst. **220**, 431 (1935).

[11] Matthew Luckiesh, *Light, Vision and Seeing* (D. Van Nostrand Company, Inc., New York, 1944), p. 172.

[12] Matthew Luckiesh, A. A. Eastman, and S. K. Guth, "Technique of using the Luckiesh-Moss visibility meter," Illum. Eng. **43**, 223, (1948).

involves the combined effects upon visibility of the contrast of the critical details with their background, the brightness level, the time available for seeing, and even the eyesight involved.

VISIBILITY LEVELS

If we look at the front page of a newspaper, we recognize many letters varying greatly in visual size. All that are visible are not equally visible. Some details are invisible. The borderline between the worlds of visibility and invisibility is threshold visibility. A visual object is near the threshold when we can barely recognize it. A visual task is near threshold visibility when we can barely see to perform it. Printed matter is near threshold when we can barely read it. Certainly none of us is satisfied with being able barely to see an object or barely to see to perform a task. Furthermore, we have as much right to have "factors of safety" in seeing as in bridge design. A reasonable factor of safety or supra-threshold visibility level is desirable from various viewpoints such as certainty and ease of seeing as well as the degree of subnormal vision that might be involved. The visibility of a visual object or task depends upon the characteristics of the object or task, upon the eyesight involved, upon the footcandle level, and upon the seeing conditions in general which involve the brightness of the immediate surroundings and the brightnesses and brightness ratios in the entire visual field. In general, satisfactory seeing and seeing conditions are largely a matter of *brightness engineering*.

The author and his colleagues have published many papers dealing with supra-threshold visibility levels, so there is little need of discussing the matter in detail here. However, Simonson and Brozek state that Luckiesh and Moss "recommend as much as 500 to 3000 footcandles for certain jobs requiring discrimination of fine details." This is an example of carelessness that leads to misunderstanding. Luckiesh and Moss did not "recommend" these levels but they have shown that illumination levels up to 3000 footcandles are necessary to raise the visibility level of certain tasks to the same visibility level as 8-point type of excellent printing with black ink on white paper when it is illuminated to a level of 10 footcandles.[13] The recommendations of Luckiesh and Moss for these tasks of inherently low visibility are merely stated as "100 footcandles or more" and they stated that their recommendations "were the first specifications of minimum desirable footcandles which attempted to recognize and to harmonize economic, practical, hygienic, and humanitarian aspects." (See reference 13, p. 346.)

EASE OF SEEING

For many years the author and his colleagues have prosecuted a continuous series of researches along the avenue of ease of seeing. This is an aspect quite distinct from other major aspects such as the relationships of light to the limitations and abilities of the visual sense, the influence of light on visibility and supra-threshold visibility, and the effects of light and lighting upon conservation and achievement. One of the major objectives has been to establish *ideal* footcandle levels for specific tasks, and particularly for reading matter which is inherently of relatively high visibility compared with many other tasks of critical seeing. An over-all objective of this line of research also has been to establish the optimum brightness level, under specified conditions, for maximal ease of seeing. Therefore, there is no justification in the statement by Simonson and Brozek that:

> "From Luckiesh and Moss' discussion it may be inferred that the desirable levels exceed the recommended levels by at least two or three times and no hint has been given that high illumination-levels might produce a deterioration of performance, even when glare is excluded."

If they had read sufficiently and carefully, they would have found such statements by Luckiesh and Moss (reference 13, p. 51) as this:

> "It should be pointed out that a brightness which is quite tolerable when critical seeing is not being done may become intolerable when one is engaged in critical seeing. However, this limiting brightness has not been established; it appears to be well above the brightness produced by 100 footcandles on a diffusely reflecting white paper. This means it is well above 100 millilamberts or footlamberts."

This statement is also based upon the results of many extensive researches and is generally

[13] Matthew Luckiesh and F. K. Moss, *The Science of Seeing* (D. Van Nostrand Company, Inc., New York, 1938), p. 343.

supported by the relatively restricted results reported by Simonson and Brozek and others. The brightness of a green grass lawn under 8500 footcandles of midsummer sunlight and skylight is commonly as high as 400 foot-lamberts, which is also the brightness of a deep blue sky. Even a sunlit distant woods may have a brightness of 200 foot-lamberts. The brightness level of a textile having a diffuse reflectance of 10 percent has a brightness of only 100 foot-lamberts when illuminated with 1000 footcandles. Certainly from this viewpoint the garment worker sewing with black thread on dark goods could be provided with 1000 footcandles on his work without exceeding a conservative limit of comfortable brightness levels. Obviously, various factors of the surroundings must also be given attention. In fact, providing good seeing conditions, including reasonable visibility levels, is largely a matter of brightness engineering.

For a decade preceding World War II, the author and his colleagues intensively studied the rate of involuntary blinking as a possible criterion of ease of seeing. More than 40 different researches were prosecuted in the study of a number of factors which influence visibility or ease of seeing or both. After the initial investigations, our technique always included determinations of the blink rate *while reading*. This important fact was overlooked by some later investigators. Furthermore, some neglected to take into account other factors which influence the blink rate. Actually, in the few cases where the blink rate was studied under conditions comparable with those reported by the author and his colleagues, there is no serious disagreement as is adequately discussed elsewhere.[14–18] Even Simonson and Brozek confirm our results insofar as the techniques are comparable. Their lack of care and discrimination in their criticism includes their comments on the results of our extensive investigations of nervous muscular tension[19] and heart rate.[16,20,21] No one as yet has published any contradictory results obtained by comparable techniques.

Apparently, Simonson and Brozek made their measurements of phorias some time after the subjects completed their work test. Luckiesh and Moss[22] obtained a smaller decrease in "convergence reserve" immediately after the subjects had read for an hour under 100 footcandles than after reading for an hour under one footcandle. However, we found and clearly stated[22] that this effect disappears in about 10 minutes after the cessation of reading.

NORMAL AND SUBNORMAL VISION

Subjects with average normal vision are commonly used in researches in the realms of light, vision, and seeing. In considering these results in terms of lighting practice, for example, the fact that the average vision of any large group is generally less than so-called average normal vision generally has been ignored. In other words, the large percentage of persons whose vision is, to some degree, subnormal are overlooked. Subnormal vision is generally aided by increasing the footcandle level over an extensive range just as normal vision is aided. However, the important fact is that footcandle levels generally must be considerably greater if a given object or task is to be at the same visibility level for persons with subnormal vision as for persons with so-called normal vision. The author and his colleagues have devoted considerable study (see reference 11, Chapters II and XII) to this aspect and insist that subnormal vision be taken into account in considerations of footcandle levels. The results of such studies show that large

[14] Matthew Luckiesh, S. K. Guth, and A. A. Eastman, "The blink-rate and ease of seeing," Illum. Eng. **42**, 584 (1947).

[15] Matthew Luckiesh, "Reading and the rate of blinking," J. Exper. Psychol. **37**, 266 (1947).

[16] Matthew Luckiesh, "Comments on criteria of ease of reading," J. Exper. Psychol. **36**, 180 (1946).

[17] Matthew Luckiesh, "On the rate of involuntary blinking," Illum. Eng. **39**, 69 (1944).

[18] Matthew Luckiesh and F. K. Moss, *Reading as a Visual Task* (D. Van Nostrand Company, Inc., New York, 1942).

[19] Matthew Luckiesh and F. K. Moss, "A correlation between illumination intensity and nervous muscular tension resulting from visual effort," J. Exper. Psychol. **16**, 540 (1933).

[20] Matthew Luckiesh and F. K. Moss, "Criteria of readability," J. Exper. Psychol. **27**, 256 (1940).

[21] Matthew Luckiesh and F. K. Moss, "The effect of visual effort upon the heart-rate," J. Gen. Psychol. **13**, 131 (1935).

[22] Matthew Luckiesh and F. K. Moss, "Fatigue of convergence induced by reading as a function of illumination intensity," Am. J. Ophthal. **18**, 319 (1935). Matthew Luckiesh and F. K. Moss, "Fatigue of the extrinsic ocular muscles while reading under sodium and tungsten light," J. Opt. Soc. Am. **25**, 216 (1935). Matthew Luckiesh and F. K. Moss, *The Science of Seeing* (D. Van Nostrand Company, Inc., New York, 1937), p. 234.

factors of safety in the specification of footcandles are needed for this reason alone.

The increase in eye-defectiveness from grade to grade among school children and from year to year among adults should receive more attention than it does. Obviously, it is difficult to determine how much of this might be due to the performance of visual tasks of inherently low visibility under relatively poor lighting and seeing conditions which have commonly prevailed. However, to an open-minded person familiar with the vast amount of available data on eye-defectiveness and with the relationships of light, vision, and seeing, there is adequate encouragement that higher footcandle levels, properly obtained, might be helpful, to some degree at least, in reducing the degree and prevalence of eye-defectiveness. For example, an authoritative survey[23] showed that 78 percent of the garment workers had defective vision. Many of their tasks are certainly of inherently low visibility, and footcandle levels provided for these tasks have not taken this into account. On the simple basis of visibility level, the footcandle levels provided for these tasks should be many times those provided for office workers, for example, if the garment workers are to be comparably aided by light. These are mere glimpses into an aspect of light, vision, and seeing which is worthy of a great deal of attention instead of being largely ignored.

RECOMMENDED FOOTCANDLE LEVELS

There are four distinct types of footcandle levels which, in a recent paper,[8] the author defined and discussed. For the present purpose they need be discussed only briefly in order to understand that recommended footcandle levels, in general, are compromises of many factors including, too often, compromises between knowledge and lack of knowledge. In considering any *one* visual task being performed by any *one* individual, there are four different footcandle levels. Three of these—threshold, optimum, and ideal—are determinate. This is not true of a *recommended* footcandle level unless it coincides with a proved *ideal* level or unless one is satisfied with an *optimum* level determined by a restricted criterion.

The threshold footcandle level is that at which specific individuals can barely perform a specific task of seeing. The threshold level varies greatly with the task and eyesight involved. For such a task as reading, one can easily determine the footcandle level at which he can barely read the material. The threshold level for any other task is determinable, but often it is not so obvious. This level is the only one of the four types that is determinable by simply trying to perform the task. Without some device such as a visibility meter, one cannot determine how far a given footcandle level is above the threshold level in a given case.

Ideal footcandle levels may be defined as those which make it possible for specific individuals to perform specific tasks with maximum ease or with a minimum expenditure of human resources. A limiting factor is the brightness level of the task. Certainly midday midsummer sunlight produces a brightness level of the white page of printed matter too great for comfortable prolonged reading. Apparently the ideal brightness level for prolonged critical seeing is above 100 foot-lamberts. The author has estimated from the limited data obtained at higher brightness levels that the optimum brightness level may be three or four times that value.

The use and meaning of the word *optimum* are sometimes confusing but it is generally applied to a specific variable, criterion, or result. For example, the rate of performance of a given task, which depends upon vision, increases as the footcandle level is increased up to a certain value. When the rate of performance reaches an optimum, that level of illumination is said to be the optimum footcandle level. In the early stage of development of lighting practice, the rate of production or performance of useful work was given a great deal of attention. However, the science of seeing has revealed many other important aspects so that at the present time performance rate no longer dominates lighting practice.

The only person who prominently advocates this criterion widely and almost exclusively is M. A. Tinker, who has depended upon his determinations of the rate of reading for very brief

[23] Leo L. Mayer, "Eyesight in industry," Arch. Ophthal. 27,,375 (1942).

periods. Tinker does not appear to realize that the optimum rate of reading is determined by such factors as reaction times, familiarity, intelligence, habits, etc. The author and his colleagues have shown that the rate of reading is a relatively insensitive criterion and that it is a grossly inadequate criterion upon which to base recommendations of footcandle levels. The difference in the breadth of our viewpoint and in the scope of our researches and those of Tinker is illustrated by comparing the works indicated in references 18 and 24. Obviously, the rate of reading is not a criterion of the optimum or desirable footcandle level for the garment worker, for example, who sews with black thread on textiles of low reflectance. Likewise, it is not a criterion of the optimum footcandle level for myriad other tasks. Such a criterion is unrevealing of other factors such as visibility level and ease of seeing. Furthermore, there are other practicable and more appropriate criteria for indicating desirable footcandle levels, so that it remains a mystery why anyone continues to depend upon very limited and restricted results and to criticize the formidable amount of knowledge that is now available from extensive researches and far more comprehensive criteria and techniques.

With these glimpses of three types of footcandle levels that are determinable, we come to the matter of recommended footcandle levels. Generally these are steps away from threshold levels and toward ideal levels. Usually they are compromises of many factors, including monetary cost and practicability. They are increased as knowledge of the factors and effects of seeing increases and as the efficiency of light production increases. Recommended footcandle levels do not remain fixed and will not until they equal the ideal levels. Considering all the knowledge available, recommended footcandle levels at present are generally conservative and are ultraconservative for prolonged critical seeing involving tasks of inherently low reflectance and low visibility. They are still more conservative when applied to prolonged critical seeing by persons with subnormal vision. Over-all, the footcandle is an inadequate and misleading term. Brightness level is more directly important than footcandle level, and visibility level is more directly important than brightness level. If we have to live with footcandle levels, let us think of them as merely a means to an end.

Actually, in the work-world where serious attention has been given to footcandle levels and other factors which influence seeing conditions, footcandle levels are often higher than the current recommendations of the Illuminating Engineering Society. This tendency appears to be more or less parallel to the growth of the science of seeing which reveals the relationships of light, lighting, brightness, and color to quick, certain, and easy seeing. The efficiency of the human seeing-machine is determined not only by the useful work he does, but also by the waste of his resources in useless work. This extended concept is one of the important contributions of researches in the realm of light, vision, and seeing. At present, recommendations of footcandle levels are, for the most part, only vaguely influenced by the important concepts of ease of seeing and supra-threshold visibility level. They are intimately related, and the ideal footcandle level determined by measurements of visibility level in any case is limited by the optimum brightness level of the task, which apparently is at least 100 foot-lamberts as determined by the combined results of researches in the realm of ease of seeing. If these two concepts were adequately taken into account, recommended levels of illumination for many tasks of prolonged critical seeing will not be merely "100 footcandles or more." In many cases they will be hundreds of footcandles. Even 1000 footcandles or more will not be confined largely to surgeries as they are at present. These relatively high levels of illumination can be obtained now by adding direct or supplementary light to general illumination. There is much evidence from the viewpoint of seeing and of seeing conditions that this is a practical and desirable means of providing adequate light and proper lighting for any specific tasks requiring relatively high footcandle levels.

[24] D. G. Paterson and M. A. Tinker, *How to Make Type Readable* (Harper and Brothers, New York, 1940).

An Infra-Red Pupillometer

M. KOOMEN, R. TOUSEY, AND H. A. KNOLL*
U. S. Naval Research Laboratory, Washington 20, D. C.

(Received April 2, 1948)

A pupillometer for measuring pupil diameters in total visual darkness has been constructed. It consists of an infra-red electron telescope arranged to observe the eye from a distance of about six inches. An attached light source irradiates the eye with invisible infra-red radiation. A scale is projected into the field of the instrument to measure the pupil diameter directly. The instrument can be used in dim and bright visible light as well as in darkness. Other possible uses of the instrument for infra-red examination of the eye are suggested.

AS part of a program of study of night vision, it was desired to measure the pupil diameter of the subject's eyes when he was adapted to total darkness and to fields of low brightness. Various methods were available, but none was completely satisfactory. Instruments for measuring the pupil entoptically have been described by Jones,[1] Blakesley,[2] and Moss.[3] Although useful in very dim light, they are not suitable for work in total darkness. Furthermore, the methods require training on the part of the subject. They suffer also from the fact that parts of the instruments are apt to obstruct a portion of the adapting field. Flash photography[4] is a method that has been used with good results. If an infra-red flash were employed, the effect upon adaptation would be small. The principal disadvantage in the photographic method, aside from the effect on adaptation and the delay in obtaining the results, is the error introduced by pupil fluctuations. Since the pupil ordinarily fluctuates constantly, it would be necessary to take a series of photographs in order to obtain a correct average pupil diameter.

In the early 1930's there were developed electronic image tubes which converted an image produced by infra-red radiation into a visible image. In recent years such devices have been perfected and produced commercially. The RCA 1-P-25 tube and its use in infra-red telescopes have been described by Morton and Flory.[5] It was evident that an infra-red telescope could be modified to observe the pupil and measure the pupil diameter in total visual darkness and also, of course, under conditions of dim illumination. Such a device would be free from most of the difficulties mentioned above. In addition, the pupil could be observed continuously, pupil fluctuations could be observed and their magnitude measured, and reactions of the pupil to various stimuli could be studied as they occurred.

A pupillometer employing an infra-red telescope was constructed according to the plan shown in Fig. 1. The essential parts of the instrument were (1) a system for irradiating the eye with infra-red radiation, (2) an infra-red telescope arranged to focus upon an object six inches from the front of its objective lens, and (3) a means of projecting a scale into the telescope field.

The eye of the subject was irradiated with infra-red radiation by means of a 50-watt projection lamp and a 25-diopter f:1 lens which formed an out-of-focus image of the lamp filament upon the subject's eye. The lamp and the lens were enclosed in a light-tight lamphouse, as shown in the figure. The aperture of the lamphouse was covered with a filter which transmitted as much infra-red as possible and absorbed the visible radiation. Polaroid infra-red filters XR7X25 or XR5X10, described by Blout and Associates,[6] were found to be satisfactory for the purpose. A single thickness of XR5X10 material was usually sufficient to prevent transmission of visible light, while a double layer of XR7X25 material was necessary. However, some subjects were still able to see a little deep red light

* Now at the Ohio State University.
[1] R. C. Jones, J. Opt. Soc. Am. **29**, 531 (1939).
[2] T. H. Blakesley, Phil. Mag. **20**, 966 (1910).
[3] F. K. Moss, J. Opt. Soc. Am. **22**, 735 (1932).
[4] H. S. Coleman, J. Opt. Soc. Am. **37**, 434 (1947).
[5] G. A. Morton and L. E. Flory, RCA Rev. **7**, 385 (1946).
[6] E. R. Blout, W. F. Amon, Jr., R. G. Shepherd, Jr., A. Thomas, C. D. West, and E. H. Land, J. Opt. Soc. Am. **36**, 460 (1946).

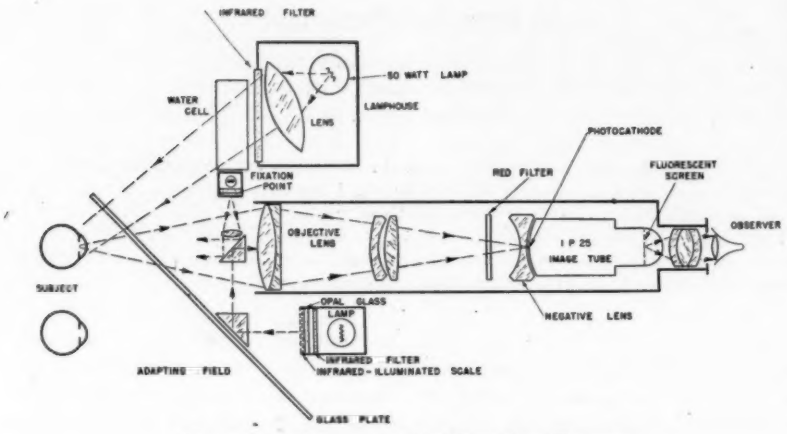


FIG. 1. Optical diagram of the infra-red pupillometer.

when they looked at the light source foveally. We found it desirable, therefore, to use a double layer of XR5X10 material at all times. Some of the infra-red radiation was, of course, sacrificed but the amount transmitted was entirely adequate. Corning glass No. 2540 was also found to be a useable filter. It was found necessary to use thicknesses greater than 5 mm to avoid transmission of perceptible amounts of visible light. The far infra-red was filtered out by a water cell about 2 cm thick. The subject therefore felt no heat. It was calculated from the spectral energy distribution of the light source, the spectral transmission of the filters and watercell, and the spectral sensitivity of the image tube that the wave-length band used by the instrument lay between 0.92 and 1.2μ, with a peak at 1.02μ. No indication was found that this radiation caused any pupil reflex.

The image-forming apparatus consisted of an objective lens, an RCA 1-P-25 image tube, and an eyepiece. The objective was a Bausch and Lomb Cinephor lens of 3.5-inch focal length and f:2.1 aperture. This projected an infra-red image of the subject's eye, at approximately unit magnification, upon the photo-cathode of the image tube. Since the surface of the photo-cathode was curved, the image was made to conform to this curvature by placing a strong negative field-correcting lens very close to the photo-cathode. The photo-cathode consisted of a caesium silver oxide film deposited upon the inner surface of the curved glass end of the image tube. Emission of electrons took place from the film in proportion to the intensity of the various parts of the infra-red image incident upon it; a system of electron lenses accelerated these electrons and formed an inverted electron image upon the fluorescent screen at the other end of the tube. The final electron image was demagnified about two times. The fluorescent screen, in turn, emitted light in proportion to the intensity of electron bombardment, and the original infra-red image was therefore reproduced in visible light. The visible image was viewed through an ocular consisting of the hemispherical glass end of the image tube and a 9-power Hastings triplet. The total magnifying power of the ocular was about 15, and the magnifying power of the entire instrument, considered as a microscope, was $1\times\frac{1}{2}\times15=7.5$. Since the image tube re-inverted the image, the final image seen by the observer was erect.

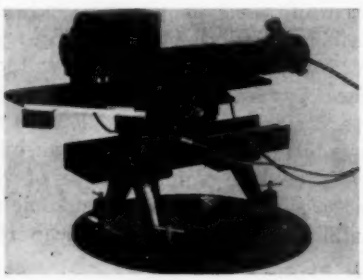


FIG. 2. The infra-red pupillometer.

Because the photo-cathode was semitransparent, the subject could ordinarily see the green glow of the fluorescent screen when he looked into the objective lens. To prevent this green light from returning to the subject's eye, a red filter (Wratten No. 29) was placed between the photo-cathode and the objective.

The scale against which the pupil diameter was measured was reflected into the objective by the two small prisms shown in Fig. 1. The scale was produced photographically and consisted of transparent lines 0.1 mm wide and 1.0 mm apart. It was backed by opal glass and illuminated from behind with a small 3-candle-power lamp. The visible light was removed by a thin piece of Corning 2540 glass. The observer conducting the measurement therefore saw a bright image of the scale superimposed upon the subject's eye. Since both the scale and the eye pupil were focused upon the photo-cathode by means of the same objective lens, both images were magnified equally and suffered equal amounts of distortion, and the scale indicated pupil diameter directly.

When desired, an adapting field was presented to the subject by reflection from a diagonally placed glass plate located between the subject and the instrument. To prevent the subject's eye from wandering, a dim red fixation point was introduced by reflection from the hypotenuse face of one of the small prisms, as shown in the figure. Use of the fixation point was of assistance but was not necessary. In all cases so far examined, the presence of the fixation point did not measurably affect the subject's pupil size.

Figure 2 shows a completed model of the pupillometer. The diagonal glass plate, ordinarily attached to the front of the instrument by means of the clamps shown, was removed to show the rest of the instrument more clearly. The fixation point, discussed above, was not present on the model shown. The infra-red illuminating system, telescope, and scale were all fixed with respect to one another and were moved as a unit to focus the image of the subject's pupil sharply upon the photo-cathode of the image tube. In addition to a back-and-forth motion along the axis of the telescope, a short lateral motion was provided, so that the pupil image could be brought to a convenient part of the scale image. Limited tilt and tip motions were provided by screws in the three legs of the instrument. The subject's head was supported in a rest. The power supply for the image tube and the rheostat controlling the intensities of the projection and scale lamps were installed in a small control box, not shown in the figure.

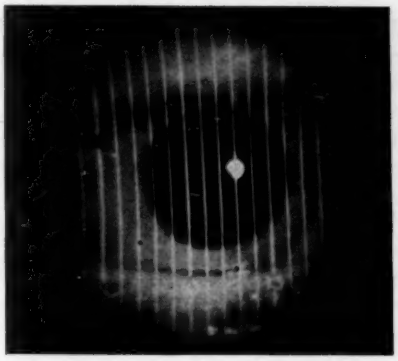


FIG. 3. Photograph of the left eye of subject *A*, taken through the eyepiece of the infra-red pupillometer.

Figures 3 and 4 are photographs of the left eyes of two subjects, taken through the eyepiece of the pupillometer. The vertical lines are the scale image, projected into the instrument. The pupil is the black circular area, and the iris and sclera appear almost uniformly bright. The contrast between iris and sclera depended upon the direction from which the infra-red irradiation came; the iris became darker and the contrast with the sclera greater when the irradiation was made more oblique. The sclera always appeared

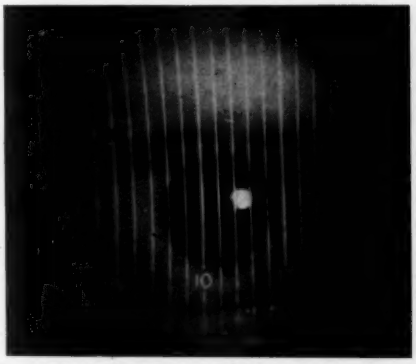


FIG. 4. Photograph of the left eye of subject *B*, taken through the eyepiece of the infra-red pupillometer.

bright. There was very little difference in appearance between brown and blue irides under the described conditions of irradiation. Brown irides appeared slightly lighter than blue.

The eyes shown in Figs. 3 and 4 had been in complete darkness for about five minutes and remained in complete darkness, except for the dim red fixation point, while the photographs were taken.

The pupil of subject *A* (Fig. 3) was 8.8 mm in diameter and very nearly circular. The pupil of subject *B* (Fig. 4) was about 6.7 mm in diameter. It was noticeably non-circular and somewhat serrated near its upper edge. The bright spot near the center of each pupil was the familiar "highlight" and was a small out-of-focus image of the projection lamp filament formed by the cornea.

The accuracy of pupil measurements made with the instrument was determined principally by the quality of the image tube. The tube used was capable of resolving a high contrast image of 15 lines per mm near the center of the photocathode. An occasional tube may be capable of resolving a little more than this. With this tube it was found possible to measure pupil diameters to approximately 0.1 mm. If greater accuracy were desired, the magnification of the objective, together with the infra-red illumination on the eye, could be increased. The scale lines could also be made somewhat narrower and closer together. However, pupil fluctuations appeared to set a practical limit to the accuracy worth trying to obtain, and 0.1 mm is well within this limit.

The luminescence of the fluorescent screen of the tube did not decay instantly when the infra-red was cut off. At the light levels employed the decay time to ten percent was of the order of 50 milliseconds. This response was sufficiently rapid, however, to permit studying fluctuations of the pupil and to observe ordinary eye movements.

In addition to the measurement of pupil diameter there are various other possible uses for the instrument. The 1-P-25 tube has its sensitivity limit at 1.2μ, while the best infra-red-sensitized photographic emulsion has a limit at about 1.35μ. The infra-red pupillometer, therefore, should give results comparable with infra-red photography. A number of applications of infra-red photography have been given by Clark.[7] Concerning infra-red examination of the eye, Clark cites the work of Dekking[8] and Mann[9] who were able to observe defects in the iris, either congenital or operative, through a dense corneal opacity. We have attempted to do this with the infra-red pupillometer. One person, blind in both eyes because of dense corneal opacities, was available. The irides or pupils could not be seen by ordinary visual inspection of the eyes. Through the infra-red pupillometer a central dark spot of irregular pattern was seen and was believed to be the pupil. No other detail, however, was visible.

The pupillometer can be easily modified to measure the convergence angle between the subject's eyes and to observe possible nystagmus or strabismus, when the subject is in total darkness. To do this the aperture of the objective could be split by means of a pair of prisms or mirrors so that the two halves of the objective would view the two eyes of the subject, respectively. Separate infra-red beams on the two eyes would probably be necessary. The observer would see in the field of the instrument the two eye pupils of the subject, either superimposed or separated by a convenient distance, depending upon the orientation of the mirrors. The apparent separation of the images of the two pupils would then depend upon the convergence of the subject's eyes, and the millimeter scale in the instrument could be calibrated in terms of convergence. This same calibration could be accomplished in terms of the positions of the images of the highlights in the subject's eyes.

[7] W. Clark, *Photography by Infrared* (John Wiley and Sons, Inc., New York, 1939).
[8] H. M. Dekking, Graefe's Archiv. f. Ophthal. **130**, 373 (1933). *ibid.* **133**, 20 (1934), Part II.
[9] W. A. Mann, Arch. Ophthal. **13**, 985 (1935).

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 8 AUGUST, 1948

Distortion of Stereoscopic Spatial Localization

KENNETH N. OGLE
Division of Physics and Biophysical Research, Mayo Clinic and Mayo Foundation, Rochester, Minnesota,

AND

PAUL BOEDER,
Bureau of Visual Science, American Optical Company, Southbridge, Massachusetts

(Received January 14, 1948)

The apparent distortion of stereoscopic space perception, caused by unequal magnifications of the images of the two eyes, as described by Ames, is studied from a mathematical point of view. It is shown that, at least qualitatively, the spatial distortions can be predicted on a geometric basis.

INTRODUCTION

A CHANGE in the magnification of the retinal image of one eye, caused, for example, by an appropriate lens placed before the eye, changes the disparities between the images of the two eyes of all objects in space that are seen binocularly. Accordingly, that sense of space derived from stereoscopic vision is affected, in that the sizes, shapes, and distances of objects appear distorted. This phenomenon is readily observed in surroundings where the stereoscopic stimuli outweigh the perspective and other empirical clues to space localization.

Especially striking is the apparent distortion of space caused by a *meridional* "size" lens, that is, an afocal lens the magnification of which is effective in one meridian only.*[1] A cylindrical ophthalmic lens used for correcting astigmatism will also introduce a meridional magnification of the image and will cause the same types of distortion. Changes in the apparent position and forms of objects due to certain ophthalmic lenses had of course been observed many years ago and have been described by Wadsworth,[2] Culbertson,[3] and Lippincott.[4] Green[5] was the first, however, to interpret them as due to the stereoscopic perception of the changed relationship between the retinal images in the two eyes. Friedenwald,[6] Koller,[7] again later Lippincott,[8] and von Rohr[9] affirmed this explanation. Similarly, Hillebrand[10] found that at near vision a thick plate of glass before one eye caused a change in location of the apparent frontal parallel plane horopter.

More recently Ames[11] has made a systematic study of these apparent spatial distortions, their dependence on the particular changes introduced between the images of the two eyes, and their relation also to the more general factors of spatial localization. The particular distortions can be reduced to four basic types corresponding to their respective causes:

(a) The image of one eye is magnified in the horizontal meridian, for example, with a meridional "size" lens, axis 90°.**

* These afocal lenses have cylindrical surfaces with parallel axes. The percentage magnification is approximately equal to the product of the dioptric power of the surface (in diopters) and two-thirds of the thickness (in centimeters). Cf. K. N. Ogle, "The correction of aniseikonia with ophthalmic lenses," J. Opt. Soc. Am. **26**, 323 (1936).

[1] K. N. Ogle, "The correction of aniseikonia with ophthalmic lenses," J. Opt. Soc. Am. **26**, 323 (1936).

[2] O. F. Wadsworth, "On the effect of a cylindrical lens, with vertical axis placed before one eye," Tr. Am. Ophth. Soc. 342 (1876).

[3] H. Culbertson, "Binocular astigmatism," J. Am. Med. Assn. **11**, 622 (1888); Am. J. Ophth. **5**, 117 (1888).

[4] J. A. Lippincott, "On the binocular metamorphopsia produced by correcting glasses," Arch. Ophth. **18**, 18 (1889).

[5] J. Green, "On certain stereoscopical illusions evoked by prismatic and cylindrical glasses," Tr. Am. Ophth. Soc. 449 (1889).

[6] Harry Friedenwald, "Binocular metamorphopsia produced by correcting glasses," Arch. Ophth. **21**, 204 (1892).

[7] C. Koller, "Ueber eine eigenthümliche Sorte dioptrischer Bilder: Ein Beitrag zur Theorie der Cylinderlinsen," Arch. f. Ophth. **32**, 169 (1886), Part 3; "The form of retinal images in the astigmatic eye," Tr. Am. Ophth. Soc. **6**, 425 (1892).

[8] J. A. Lippincott, "On the binocular metamorphopsia produced by optical means," Arch. Ophth. **46**, 397 (1917); also J. A. Lippincott, "New tests for binocular vision," Tr. Am. Ophth. Soc. 560 (1890).

[9] M. von Rohr, "Die Brille als optische Instrument," (Verlag Julius Springer, Berlin, 1921), third edition, pp. 195.

[10] Franz Hillebrand, "Die Stabilität der Raumwerte auf der Netzhaut," Zeits. Psychol. **5**, 1 (1893).

[11] Adelbert Ames, Jr., "Binocular vision as affected by relations between uniocular stimulus-patterns in commonplace environments," Am. J. Psychol. **59**, 333 (1946).

** The specification of the axis is in accordance with ophthalmic practice.

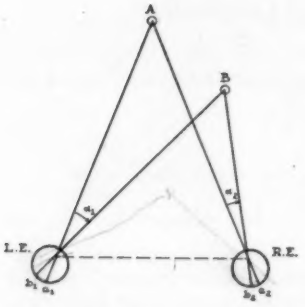


FIG. 1. The geometric disparities between images of two objects in space, as the basis for stereoscopic spatial localization.

(b) The image of one eye is magnified in the vertical meridian, for example, with a meridional "size" lens, axis 180°.

(c) The images of both eyes are magnified in oblique meridians, for example, with a meridional "size" lens, axis 45°, before the right eye and another, axis 135°, before the left eye, or *vice versa*.

(d) The images in both eyes are unsymmetrically magnified by ophthalmic prisms placed before the eyes bases-in or bases-out.

As suggested in Ames's paper, the stereoscopic spatial distortions resulting from (a) and (c), and to a certain extent from (d), may have geometric explanations. The purpose of this paper is to show that these types of spatial distortions can be predicted on a geometric basis.

That the distortions are caused only by the changes in the disparities between the images of the two eyes is clear because the small magnifications involved (those used are less than 5 percent) have little or no influence on the spatial localization derived from monocular clues, such as linear perspective, overlay, and parallax, but the keenness of our stereoscopic acuity enables us to recognize changes produced by magnifications of as low as 0.12–0.25 percent.[12–15] Since the magnifications introduced by lenses change the disparities in an orderly fashion, a mathematical analysis of the effects is possible. Only in case (b), that of magnification in the vertical meridian, a direct analysis fails because vertical disparities themselves do not give rise to a stereoscopic experience.

Suppose the two points A and B are seen binocularly, as illustrated in Fig. 1. The distance between the two retinal images (a_1b_1) of the left eye is greater than that (a_2b_2) of the right eye. The object B is perceived nearer than A through the process of stereopsis, which depends only on the subconscious response by the higher centers of the cortex to those differences in retinal distances. The *geometric* disparity between the two images in the two eyes can be described in terms of the angles subtended by the two objects at the two eyes, $\Delta = \alpha_1 - \alpha_2$, where $\alpha_1 = \sphericalangle ALB$ and $\alpha_2 = \sphericalangle ARB$ are positive if measured in a clockwise direction. If Δ is positive, B appears nearer than A. These angles themselves (retinal dimensions) do not provide sufficient means for the judgment of the *absolute* distance of either A or B from the eyes, but their *difference* does provide the means for the *relative* localization of the two points. Stereopsis can occur only, of course, if there are at least two points in the visual field, for otherwise disparities do not exist. In a binocular visual field containing many object points, a geometric disparity exists for any two, but usually the disparities are referred to the images of the point of fixation (the *Kernpunkt* of visual space).

If the image of one eye is magnified, it is clear

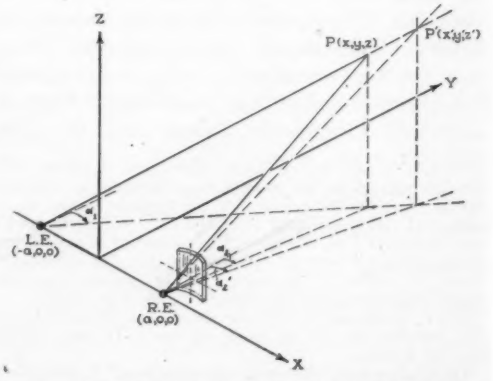


FIG. 2. The effect of a meridional magnifying lens on the apparent positions of objects in the binocular visual field.

[12] Adelbert Ames, Jr. and K. N. Ogle, "Size and shape of ocular images: III. Visual sensitivity to differences in the relative sizes of the ocular images of the two eyes," Arch. Ophth. **7**, 904 (1932).

[13] Adelbert Ames, Jr., "Aniseikonia—a factor in the functioning of vision," Am. J. Ophth. **18**, 1014 (1935).

[14] K. N. Ogle, "Association between aniseikonia and anomalous binocular space perception," Arch. Ophth. **30**, 54 (1943).

[15] K. N. Ogle and V. J. Ellerbrock, "Stereoscopic sensitivity in the space eikonometer," Arch. Ophth. **34**, 303 (1945).

that the retinal distance between the images of A and B for that eye is increased. This changes the value of Δ, and, corresponding to the new disparity, the points A and B appear differently oriented. The essential problem of this paper is to find the new relative positions of A and B, such as would correspond to the new disparity.

I. A MAGNIFICATION OF THE IMAGE OF ONE EYE IN THE HORIZONTAL MERIDIAN (A LENS AT AXIS 90°)

The characteristic apparent distortion of space caused by a magnification in the horizontal meridian of the image (say) of the *right* eye by 3 percent can be described (cf. Ames's paper[11]) as follows:

(a) The level floor (grass) on which the individual stands appears like a side hill, that is, tipped down at the right, up at the left.

(b) A ceiling appears tipped in the opposite direction; higher on the right, lower on the left.

(c) Walls on the left side appear to be more or less parallel to themselves but nearer, with a corresponding apparent diminution in the size of the details on the wall.

(d) Walls on the right side appear to be more or less parallel to themselves but much farther away, with a corresponding apparent enlargement of the details on the walls.

(e) Front walls appear more or less vertical but rotated considerably, with the right side farther away, the left side nearer. The right side of objects appears enlarged, the left side shrunken.

This characteristic distortion can be predicted, at least qualitatively, by means of the following simple analysis.

In Fig. 2, let the "centers" of the left and right eyes (assuming coincidence of the entrance pupils with the centers of rotation) be located in a rectangular coordinate system at points $(-a, 0, 0)$ and $(a, 0, 0)$, respectively. The XY plane is the so-called objective visual plane, XZ, the "frontal plane" and YZ the "median plane" of the eyes. Consider a point at a very great distance on the Y axis, a second point P at x, y, z, and their horizontal visual angles α_1 and

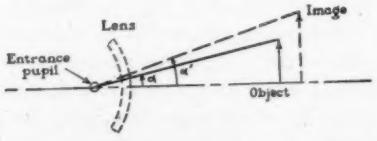


FIG. 3. Illustrating the definition of angular magnification.

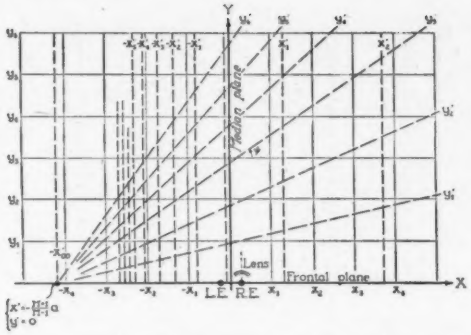


FIG. 4. Relationship between objective localization of objects in space and their stereoscopic localization when the image of the right eye is magnified in the horizontal meridian. Objective planes parallel to median plane and planes parallel to the frontal plane.

α_2 for the left and right eyes, respectively. Now let a meridional afocal magnifying lens be placed before the right eye, with geometric axis vertical (axis 90°) so as to magnify in the horizontal meridian only. The visual angle α_2 is enlarged to α_2' and the disparity $\Delta=\alpha_1-\alpha_2$ is changed. Correspondingly, the stereoscopic stimulus is changed, as though given by a point P' at x', y', and z'. The transformation of x, y, z to x', y', z' in terms of the angular magnification of the lens and the interpupillary distance $2a$ is to be found.

For the purpose of this discussion it is convenient to define the angular magnification of the lens by $M=\tan\alpha'/\tan\alpha$, as illustrated in Fig. 3. Also M is assumed constant for all values of α, though these afocal lenses usually have a small pincushion distortion. The small changes in magnification caused by the object distance will also be neglected. The lens is placed before the observer's eyes with optic axis parallel to the Y axis. However, small variations from this position do not change the general nature of the apparent space distortion.

Since the magnification acts in the horizontal meridian only, we can write

$$\tan\alpha_1=(x+a)/y, \qquad \tan\alpha_2=(x-a)/y,$$
$$\tan\alpha_1'=(x'+a)/y' \qquad \tan\alpha_2'=(x'-a)/y',$$
$$\tan\alpha_1'=\tan\alpha_1, \quad \text{and} \quad \tan\alpha_2'=M\tan\alpha_2.$$

Eliminating the $\tan\alpha$'s from these six relations we obtain the transformation equations:

$$x'=-a[(M+1)x-(M-1)a]/[(M-1)x-(M+1)a], \quad (1)$$

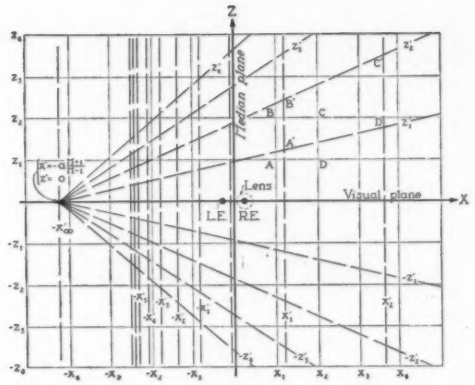


FIG. 5. Relationship between objective localization of spatial objects and their stereoscopic localization when the image of the right eye is magnified in the horizontal meridian. Objective planes parallel to visual plane and those parallel to median plane.

$$y' = -2ay/[(M-1)x-(M+1)a], \tag{2}$$

and since $z'/z = y'/y$,

$$z' = -2az/[(M-1)x-(M+1)a]. \tag{3}$$

These equations relate the coordinates of any point in the x, y, z space to one in the x', y', z' space and depend only on the magnification of the lens and the interpupillary distance. Because these transformations are linear, it follows that straight lines in one space remain straight lines in the other; conic sections in one remain conic sections in the other, and so forth. Because the lens magnifies only in the horizontal meridian, x' is independent of y and z. Thus, all planes parallel to the median plane (the YZ plane) remain parallel in the transformation, but will be displaced in the direction of the X axis. With increasing distances on the right side of the median plane (Fig. 4) the apparent displacement to the right increases and when the distance $x=a(M+1)/(M-1)$ is reached, the apparent plane should be infinitely far to the right. Planes at increasing distances on the left side appear also displaced to the right to a position never farther from the eyes than the distance $x' = -a(M+1)/(M-1)$. In fact, all planes to the left go into the region between $x' = -a(M+1)/(M-1)$ and $x' = -a(M-1)/(M+1)$. The transformed planes corresponding to equally spaced objective planes on the left have $x' = -a(M+1)/(M-1)$ as a limit plane. Planes between $x=-a$ and $x=+a$ are slightly displaced to the left.

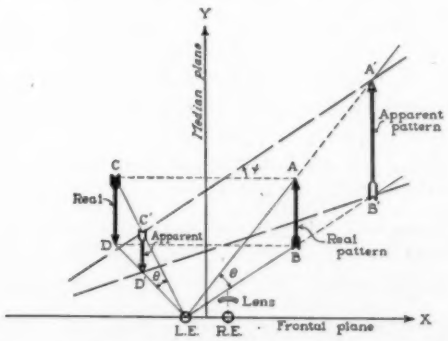


FIG. 6. The increase (or decrease) in the apparent size of patterns observed with stereoscopic vision when the image of the right eye is magnified in horizontal meridian (visual plane).

The inverse transformation equations

$$x = a[(M+1)x' + (M-1)a]/[(M-1)x' + (M+1)a], \tag{4}$$

$$y = 2aMy'/[(M-1)x' + (M+1)a], \tag{5}$$

$$z = 2aMz'/[(M-1)x' + (M+1)a], \tag{6}$$

are useful in studying the distortion of any surface or figure in space the equation of which is known.

Consider now planes parallel to the frontal plane (the XZ plane), given by $y=y_0$. The corresponding "apparent" plane is given by

$$y' = [(M-1)y_0/2aM]x' + (M+1)y_0/2M. \tag{7}$$

Each plane appears rotated about a vertical axis through the angle given by

$$\tan\psi = [(M-1)/2M][y_0/a]. \tag{8}$$

Clearly ψ increases with the distance of the objective plane and decreases with the interpupillary distance. All "apparent" planes corresponding to frontal parallel planes intersect the X axis at the point $x' = -a(M+1)/(M-1)$. These transformations and those discussed before, both illustrated in Fig. 4, conform closely with the actual appearances of side and front walls under the conditions assumed, providing those walls have few perspective features, and many details which provide stimuli for stereoscopic vision.

It remains to find the transformations affecting

floors and ceilings. Here $z=z_0$, and, as previously,

$$z'=[(M-1)z_0/2aM]x'+[(M+1)z_0/2M]. \quad (9)$$

Planes parallel to the visual (XY) plane become inclined. For planes above the visual plane (z_0 positive) the slope of the transformed planes is positive, that is, the plane is higher on the right than on the left. For those below the visual plane (z_0 negative) the transformed planes incline in the opposite direction, the right-hand side is lower than the left. These transformations are illustrated schematically in Fig. 5. Again all the transformed planes intersect the X axis at the point $x'=-a(M+1)/(M-1)$. Table I shows approximately the position of this unique point for different magnifications of the lens.

The Ames "leaf room" is a rectangular box about 7 feet (213 cm) square, open at one end, with walls vertical and the floor and ceiling level. The observer views the room from the center of the open end. To the inside surfaces of the room are stapled artificial vines with leaves individually adjusted so as to stand out more or less uniformly from the surfaces. When evenly illuminated, these leaves provide many contours which stimulate stereoscopic responses, yet introduce only a minimum of monocular perspective clues. The apparent distortion of such a room when viewed with a meridional "size" lens before the right eye is as predicted on a geometric basis. The right wall appears farther to the right, the left nearer, the ceiling is inclined up to the right, the floor down to the right.

Very striking also is the fact that in the "leaf room" the leaves on the right wall appear much larger, and those on the left smaller. Since the angular size of the leaves remains practically the same without and with the lens before the right eye, the leaf that appears farther away will also appear larger. Figure 6 illustrates the geometry of the situation. From Eqs. (2) and (3) previously given, it is clear that for a given wall

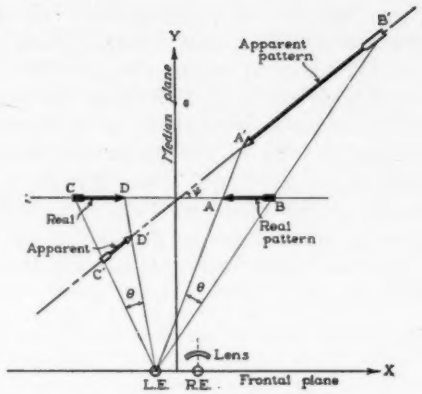


FIG. 7. The distortion of details on a frontal wall, when the image of the right eye is magnified in the horizontal meridian.

parallel to the median plane, the apparent dimensions of a figure on that wall are always proportional to the real size. That is,

$$y_2'-y_1'=k(y_2-y_1) \quad \text{and} \quad z_2'-z_1'=k(z_2-z_1),$$

where

$$k=-2a/[(M-1)x-(M+1)a];$$

there is no distortion of the figure. The natural perspective (in the sense of decrease in angular size with distance) remains the same.

This statement is not true for details on frontal walls. For these the apparent distortion changes progressively with the extent of the figure; that is, the elongation varies with x (Eq. 1), as illustrated in Fig. 7. The vertical extension of objects on a frontal wall will also be distorted, depending not only on x, but also on the distance of the figure above or below the visual plane, that is, on z (Eq. 3). The general idea of this distortion can be visualized in Fig. 5 by comparing any square in the rectangular pattern (A, B, C, D) with the transformed pattern (A', B', C', D').

One further aspect of this problem is the

TABLE I. Calculated values for the unique point on the X axis where the apparent planes of all frontoparallel planes converge for an interpupillary distance of 65 mm.

	Magnification introduced in the image of the right eye							
	1% $M=1.01$	2% $M=1.02$	3% $M=1.03$	4% $M=1.04$	5% $M=1.05$	6% $M=1.06$	7% $M=1.07$	10% $M=1.10$
x_0 (cm)	−653	−328	−220	−166	−133	−112	−96	−68

apparent rotation of a surface that is inclined with respect to the Y axis. A case in point is the Ames tippingboard, where the observer looks down on a horizontal plane, inclined with respect to the observer's visual plane, as schematically illustrated in Fig. 8. The plane can be represented by $z=(y-b)\cot i$, where b is the visual distance to the plane and i is the angle of inclination referred to the Z axis. Substituting the inverse transformation equations (4), (5), and (6), we have for the new surface

$$x'[(M-1)b]-y[2Ma]+z'[2Ma\tan i]+[(M+1)ab]=0.$$

Inspection of this equation shows that the inclination of the transformed plane is the same as before, though there is a small displacement along the Y axis because of the position of the lens. The rotation of this plane about the Z axis is the same as that for any surface at the distance b, given by (8).

The degree of rotation about the axis AA (Fig. 8) depends, however, on the inclination i; it is given by

$$\tan\psi i=[(M-1)/2M][b/a]\cos i. \qquad (10)$$

Experimental measurements of the apparent rotation of inclined surfaces when the image of one eye is magnified in the horizontal meridian are in good agreement with this relationship.[16] Therefore, this effect is often referred to as a *geometric* effect in contradistinction to the induced effect mentioned under (*B*).

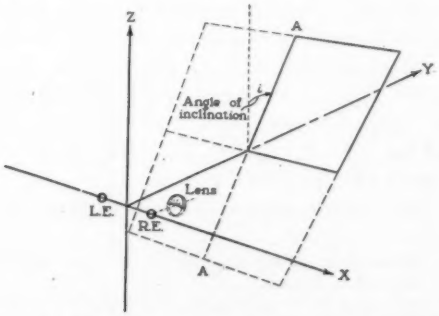


FIG. 8. The case of a surface inclined to the frontal (XZ) plane.

[16] K. N. Ogle, "Induced size effect: III. A study of the phenomenon as influenced by horizontal disparity of the fusion contours," Arch. Ophth. **22**, 613 (1939).

If the meridional lens is placed before the left eye, the direction of all the apparent distortions resulting from stereoscopic localization is reversed. ($1/M$ takes the place of M in the transformation formulas.)

Ames has shown that in surroundings with predominant monocular factors of depth perception this stereoscopic spatial distortion may be weakened. Aside from this, there are situations in which, for reasons determined by the nature of binocular space perception, the degree of the apparent distortion may not be as great as might be expected on the basis of a purely geometric analysis of the disparities introduced. For example, according to Eq. (1) the apparent distance of the right side wall of a leaf room should be infinite for an objective distance of $x=a(M+1)/(M-1)$. However, the wall, though appearing much farther away, does not appear infinitely far. Thus though other factors may enter to dampen or suppress the magnitude of the stereoscopic effect, nevertheless the stimuli must be present for an incorrect stereoscopic localization of space.

II. A MAGNIFICATION OF THE IMAGE OF ONE EYE IN THE VERTICAL MERIDIAN (A LENS AT AXIS 180°)

If the meridional "size" lens is placed before the right eye with axis horizontal (axis 180°) so as to magnify the image in the vertical meridian, the over-all localization of objects in space is again distorted. In ordinary surroundings this distortion appears identical with that resulting from the same lens placed before the other eye to magnify in the *horizontal* meridian. Thus, it seems that a vertical magnification of the image of one eye *induces* a magnification of the image of the other eye in the horizontal meridian, the effects of which have just been analyzed. The induced effect, itself presenting a new aspect of binocular vision, cannot easily be predicted on a geometric basis. It is a complex process and will not be considered further here.[17]

[17] K. N. Ogle, "Induced size effect: I. A new phenomenon in binocular space perception associated with the relative sizes of the images of the two eyes," Arch. Ophth. **20**, 604 (1938); "II. An experimental study of the phenomenon with restricted fusion stimuli," **21**, 604 (1939); "III. A study of the phenomenon as influenced by horizontal disparity of the fusion contours," **22**, 613 (1939); K. N. Ogle, "Induced size effect," J. Opt. Soc. Am. **30**, 145 (1940); K. N. Ogle, "Induced size effect with eyes in

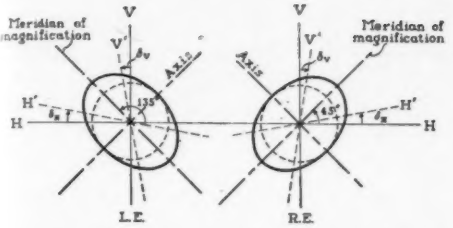


FIG. 9. The positions of the meridional magnifications at oblique axes for maximal effect (case of axes of magnifications converging up), showing the "scissors" or torsional deviations of the images of vertical and horizontal lines.

Because of the induced size effect, the magnification of the image of one eye with an "over-all" size lens (an afocal magnifying lens with spherical surfaces) causes little or no stereoscopic distortion of space. The induced effect apparently offsets the geometric effect of a horizontal magnification.

III. THE IMAGES OF BOTH EYES ARE MAGNIFIED IN OBLIQUE MERIDIANS

The third type of stereoscopic spatial distortion results when meridional magnifications are introduced before one or both eyes at oblique meridians.***,[18] A single lens before one eye that introduces a meridional magnification at an oblique axis changes the disparities of the images both horizontally and vertically and, therefore, effects stereoscopic spatial localization according to (*A*) and (*B*) described previously. In order to remove this complication, equal meridional magnifications are used symmetrically before both eyes. For the maximal effect one lens is placed at axis 45°, and the other at axis 135°. The vertical and horizontal components of the magnifications are then equal for both eyes, but there remains[19] a torsional (scissors) effect of the images which causes the characteristic spatial distortion experienced. The optical effect is illustrated in Fig. 9, where a circular object becomes elliptical owing to the meridional magnification.

Spectacles which introduce magnifications of the image of the right eye in the 45° meridian, and of the left eye in the 135° meridian (Fig. 9) cause a stereoscopic spatial distortion with the following characteristics (cf. Ames[11]):

(a) The floor appears more or less level but much nearer the observer. Details on the floor appear smaller.

(b) The ceiling appears more or less level, but much farther away, and details on it appear much larger.

(c) A frontoparallel wall appears flat but inclined, the top away from and the bottom nearer to, the observer.

(d) Side walls appear more or less flat, but with the tops leaning away from the observer.

(e) Objects appear enlarged above, shrunken below. Square objects appear trapezoidal with the top dimension longer than the bottom dimension.

If the spectacles are turned upside down, the direction of the distortion is reversed.

These distortions can also be predicted, at least qualitatively, on the basis of geometric analysis. Consider, first, the angular displacement of the image as seen through a meridional lens magnifying in an oblique meridian (Fig. 10). The magnifying effect takes place parallel to the meridian at θ. For any point ($w=\tan\alpha$, $v=\tan\beta$) the distances w_θ, v_θ, from the principal meridians of the lens are

$$w_\theta = w\cos\theta + v\sin\theta, \quad w = w_\theta\cos\theta - v_\theta\sin\theta,$$
$$v_\theta = v\cos\theta - w\sin\theta, \quad v = w_\theta\sin\theta + v_\theta\cos\theta.$$

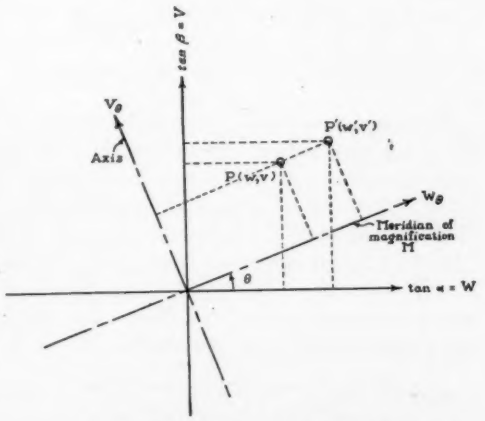


FIG. 10. Diagram for finding the displacement of the image of a point, due to a meridional magnification M at meridian θ.

asymmetric convergence," Arch. Ophth. **23**, 1023 (1940); A. S. Householder, "A theory of the induced size effect," Bull. Math. Biophys. **5**, 155 (1943).

*** For a consideration of the spatial distortions in cases of astigmatic spectacle corrections with oblique axes see: K. N. Ogle and L. F. Madigan, "Astigmatism at oblique axes and binocular stereoscopic spatial localization," Arch. Opth. **33**, 116 (1945).

[18] K. N. Ogle and L. F. Madigan, "Astigmatism at oblique axes and binocular stereoscopic spatial localization," Arch. Ophth. **33**, 116 (1945).

[19] K. N. Ogle, "Meridional magnifying lens systems in the measurement and correction of aniseikonia," J. Opt. Soc. Am. **34**, 302 (1944).

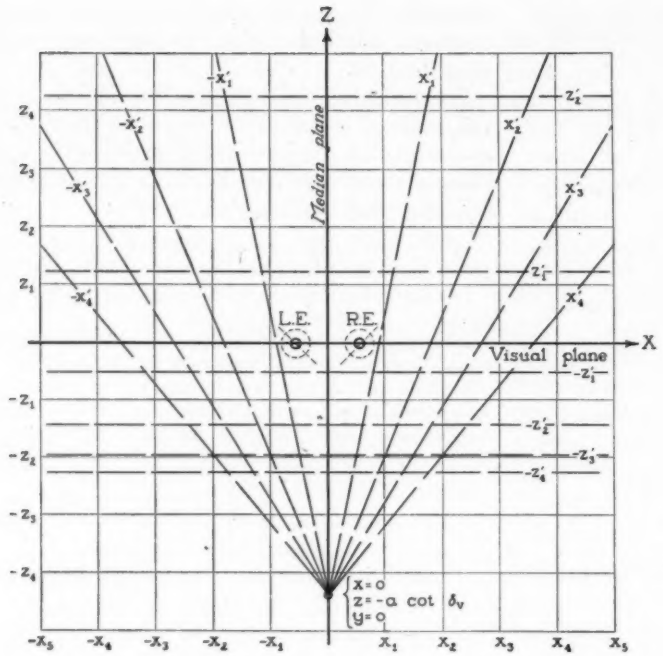


FIG. 11. The transformation of planes parallel to the median and to the visual planes caused by meridional magnification at oblique meridians.

Because of the magnification M, we have $w_\theta' = Mw_\theta$, and $v_\theta' = v_\theta$, and therefore the coordinates of the image point $P'(w', v')$, corresponding to $P(w, v)$, are

$$w' = Aw + Bv, \tag{11}$$

$$v' = Bw + Cv, \tag{12}$$

where

$$A = \tfrac{1}{2}[(M+1) + (M-1)\cos 2\theta],$$
$$B = \tfrac{1}{2}[(M-1)\sin 2\theta],$$
$$C = \tfrac{1}{2}[(M+1) - (M-1)\cos 2\theta].$$

Refer again to Fig. 2, remembering that now there are two lenses, one before the right eye with magnification M in the meridian θ, the other before the left eye with magnification M in the meridian $180° - \theta$. Using subscripts 1 and 2 for the left and right eyes, respectively, we have $A_1 = A_2 = A$, $B_1 = -B_2 = -B$, and $C_1 = C_2 = C$, and substituting point coordinates for the w's and v's in (11) and (12), we obtain

$$\begin{aligned}(x'+a)/y' &= A(x+a)/y - Bz/y,\\ (x'-a)/y' &= A(x-a)/y + Bz/y,\end{aligned} \tag{13}$$

$$\begin{aligned}z_1'/y' &= -B(x+a)/y + Cz/y,\\ z_2'/y' &= B(x-a)/y + Cz/y.\end{aligned} \tag{14}$$

In order to get a unique z', we put $z' = \frac{1}{2}(z_1' + z_2')$, an assumption that is justified because B is small. For small absolute values of x the expression represents a good approximation, and for all points on the plane $x = 0$, it is exact. Then, solving for x', y', and z', we find

$$x' = Aax/(Aa - Bz), \tag{15}$$

$$y' = ay/(Aa - Bz), \tag{16}$$

$$z' = a(Cz - Ba)/(Aa - Bz). \tag{17}$$

Both x' and y' depend on z, and thus the apparent displacement of any point changes with its position above or below the visual plane. On the other hand, z' is independent of x and y, indicating that planes parallel to the visual plane, though displaced, remain parallel in the transformation. This prediction conforms to the appearance of the floor and ceiling as experienced by observers wearing the lenses described previously. According to (17) as z becomes very large, z' approaches the value $-aC/B$. The transformed planes that correspond to equally spaced object planes approach the plane $z' = -aC/B$ as a limit. Figures 11 and 12 illustrate this transformation.

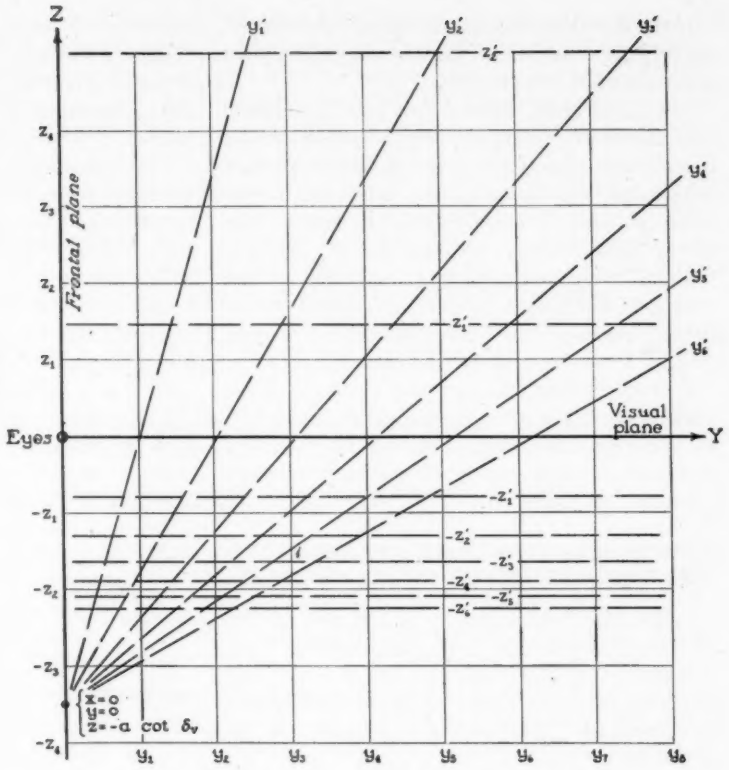


FIG. 12. The transformation of planes parallel to the frontal and to the visual planes caused by meridional magnifications at oblique meridians.

The transformation of planes parallel to the median plane ($x=x_0$) and to the frontal plane ($y=y_0$) can be determined by eliminating z from (15) and (17), and, again, from (16) and (17). We find

$$z'=[a(AC-B^2)/ABx]x'-[Ca/B], \quad (18)$$

and

$$z'=[a(AC-B^2)/By]y'-[Ca/B]. \quad (19)$$

The transformed planes corresponding to $x=x_0$ are parallel to the Y axis, but are increasingly inclined from the vertical as x_0 increases. This result conforms to the general appearance experienced by observers; side walls appear sloping away at the top, nearer at the bottom. The transformation is schematically illustrated in Fig. 11. All transformed planes pass through the point at $z=-a[C/B]$ of the Z axis.

The transformed planes corresponding to frontal parallel planes, $y=y_0$, are parallel to the X axis but are also inclined with top farther away. The angle of inclination measured from the vertical increases with increasing y_0. Again, this result describes, at least qualitatively, the appearance of a rear wall as experienced by an observer wearing the lenses described previously. Again, all transformed planes pass through the point $z=-a[C/B]$ of the Z axis. The transformation is illustrated in Fig. 12.

It can easily be shown[20] that the ratio $B/C=\tan\delta_V$, where δ_V is the rotary deviation (declination) of the images of verticals caused by the lens (cf. δV in Fig. 9). Similarly $B/A=\tan\delta_H$, where δ_H is the rotary deviation of the image of the horizontal lines. Thus the unique point in these transformations is related to the interpupillary distance and to the vertical declination angle[19,†] (cf. Fig. 13).

[20] A. Ames, Jr., K. N. Ogle, and G. H. Glidden, "Corresponding retinal points, the horopter and size and shape of ocular images," J. Opt. Soc. Am. 22, 576 (1932).

† In this connection it may be of interest to point out Helmholtz's deduction (*Physiological Optics*, the Optical Society of America, 1925, Vol. III, p. 349) of the position of the horopter surface that exists under the assumption that the meridians V' for both eyes (Fig. 13) are corresponding retinal meridians.

If the meridians of magnification are reversed, $\theta > 90°$, the direction of the distortion described previously will be reversed.

The predictions derived from this geometric analysis are in striking agreement with the actual appearances of surfaces when observed through meridional magnifying lenses. True, the limitations and conditions mentioned may render them merely qualitative, but the Ames distorted "rooms," which were designed according to the geometric distortions produced by certain magnifying lenses, appear as normal rooms if viewed through lenses having the opposite effect.

If the observer changes the position of his head, looks down or to the right or to the left, the apparent distortions change slightly, though the over-all appearance of surfaces remains essentially as predicted.

IV. THE IMAGES OF THE TWO EYES ARE MAGNIFIED UNSYMMETRICALLY IN ONE MERIDIAN

A flat ophthalmic prism placed before the eye introduces an angular magnification that increases toward the apex of the prism.[21] For practical purposes the magnification in the base-apex meridian (apex to the right) can be described by

$$M = M_0 + h \tan\alpha,$$

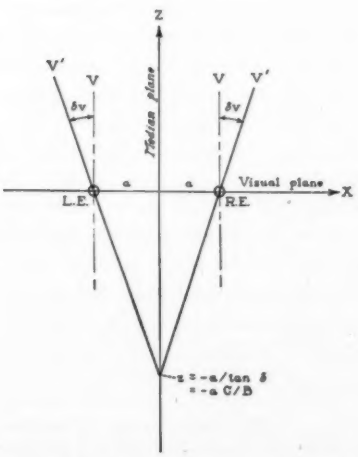


FIG. 13. Relationship between the unique point $z = -aC/B$, and the vertical declination angles of the two eyes.

[21] H. von Helmholtz, *Physiological Optics* (English translation) (The Optical Society of America, 1925), Vol. 3, p. 349.

where α, as before, is the visual angle; M_0 is a constant magnification and h is the gradient coefficient of magnification, both depending on the prism power, the base curve, and other factors.

The prismatic effect can be eliminated by placing a second prism in front of the first, base to the apex; the second prism is ground on a high base-curve (about 9-diopters base-curve), which has no gradient magnification. The combination, called by Ames an "asymmetric distortion lens," has no prismatic deviation, only a non-uniform meridional magnification.

If a pair of prisms or a pair of the prism combination lenses is placed before the eyes, with the bases of the flat prisms toward the nose (base-in) or toward the temples (base-out), the disparity of the retinal images of objects in space is changed, and consequently stereoscopic space perception is affected.

A combination lens before the left eye (base-in) changes the visual angles according to

$$\tan\alpha_1' = \tan\alpha_1(M_0 - h\tan\alpha_1),$$

and one before the right eye (base-in) according to

$$\tan\alpha_2' = \tan\alpha_2(M_0 + h\tan\alpha_2).$$

Since the constant magnification M_0 is effective equally to both eyes, we replace it, for our purpose, by unity. Introducing point coordinates, we obtain the relations

$$(x'+a)/y' = (x+a)/y - h(x+a)^2/y^2,$$
$$(x'-a)/y' = (x-a)/y + h(x-a)^2/y^2,$$

from which we find the transformation equations:

$$x' = ax[y - 2ah]/[ay - h(x^2+a^2)],$$
$$y' = ay^2/[ay - h(x^2+a^2)],$$
$$z' = ayz/[ay - h(x^2+a^2)].$$

We note that, according to this transformation, plane surfaces appear convex toward the observer. A front wall, $y = y_0$, appears cylindrically convex for $x^2 < ay_0/h$. The cross section is approximately a parabola symmetrical to $x = 0$. A side wall, $x = x_0$, $x_0 > a$, appears also cylindrically convex; for $y \gg 2ah$, the cross section is approximately a hyperbola with $x = x_0$ and $y = 2ah$, as asymptotes. The apparent convexity of floors and ceilings, $z = z_0$, is more complex.

These predictions are qualitatively borne out

by actual experience with these lenses.[11] If the lenses are placed with bases toward the temples (base-out), the apparent distortions are reversed, that is, the plane surfaces appear concave. This brief description may suffice here, for the effects of prisms are complex, not only because of additional optical distortions that are present, but also because of certain psychologic factors which, varying with the configurations and details present, affect the spatial localization.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 8 AUGUST, 1948

Reduction of Sphero-Chromatic Aberration in Catadioptric Systems*

ROBERT E. STEPHENS

(Received March 4, 1948)

Schmidt and Maksutov catadioptric systems suffer from sphero-chromatic aberrations that are of opposite sign. By the use of a hybrid system, whose corrector is a deep meniscus lens with an aspheric curve on the external surface, the sphero-chromatic residuals have been reduced and, as a consequence, usable relative apertures have been increased.

SOME time ago a Maksutov camera system of large relative aperture was designed at the National Bureau of Standards for the purpose of ascertaining whether or not such a system, employing as it does only spherical surfaces, could be used to replace a Schmidt system for certain purposes. It was found that the residual zonal spherical aberration was so large that the system was unsatisfactory. On further analysis, however, it was found that the residual sphero-chromatic aberration of the Maksutov system was opposite in sign to that of a comparable Schmidt system and of about the same magnitude. This suggested that a hybrid system, partially corrected by an aspheric surface and partially by a deep meniscus lens, would suffer substantially less from sphero-chromatic aberration than either the Schmidt or the Maksutov type. Such has indeed been found to be the case.

A hybrid Maksutov-Schmidt system consisting of a correcting plate, plane on one side, placed with its vertex at the center of the concave mirror, and a meniscus achromatic lens properly placed, could presumably be designed that would be as free from coma as the best of either of the original types. This has not been attempted by the author. Instead, several systems have been designed in which the aspheric surface is applied to the side of the meniscus lens away from the mirror.

The chromatic aberration of either a Schmidt or a Maksutov system results from the variation of correction for spherical aberration with index of refraction of the correcting plate or meniscus lens. In either of these systems the variation of spherical correction is preponderantly due to changes of the third order spherical aberration of the correcting plate or lens. In a Schmidt system the spherical aberration is perfectly corrected for one wave-length. At other wave-lengths there exists simple spherical aberration, under-corrected for longer wave-lengths and over-corrected for shorter wave-lengths. In a Maksutov system the situation is reversed. Superimposed upon the zonal spherical aberration, which is the best that can be accomplished by the meniscus lens, is over-correction of spherical aberration at longer wave-lengths and under-correction at shorter wave-lengths. Because of the large angles of incidence existing at the surfaces of the lens in this type system, aberrations of higher order are present to a greater extent than in the Schmidt system, but the differences over a small range of wave-lengths are still predominantly of the third order.

For the first trial it was intended to design a system with a relative aperture of $f/0.7$ which would be achromatic at $f/1.0$. However, because of the effect of the corrector upon the equivalent

* This work has been sponsored by the Bureau of Ships, Navy Department.

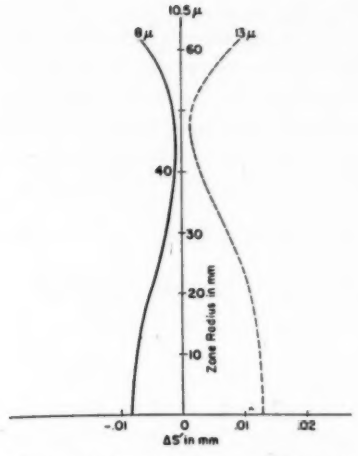


FIG. 1. Spherical aberration of first Schmidt-Maksutov hybrid. Equivalent focal length = 88.68 mm.

focal length the achromatic zone turned out to be at $f/0.89$. A first radius and a suitable thickness were assumed for the corrector. The second radius was then given the value that resulted in achromatism at the selected zone. Rays were then traced through the system at the most important wave-length and the condition of the spherical aberration determined. Different values were assumed for the first radius until a system of the Maksutov type was found in which the marginal longitudinal spherical aberration was about half that for the uncorrected mirror.

The first surface was then replaced by an aspheric surface[1] such that the optical path

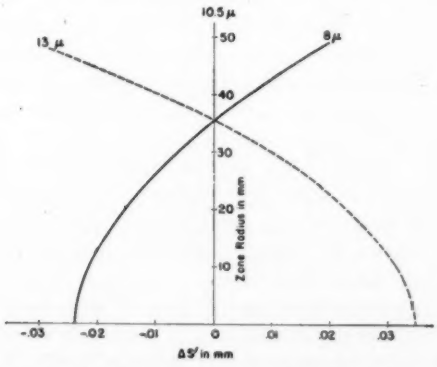


FIG. 2. Spherical aberration curves for a Schmidt camera system. Equivalent focal length = 83.56 mm.

[1] It is not the purpose of this paper to describe the methods used to compute aspheric surfaces. The procedures used were similar to the exact method of F. A. Lucy, J. Opt. Soc. Am. **30**, 251 (1940); and, for finite object distances, H. S. Friedman, J. Opt. Soc. Am. **37**, 480 (1947).

lengths for the principal wave-length from all points of an incident plane wave to the chosen focus were made equal. The chosen focus was that corresponding to the ray of the principal wave-length through the achromatic zone of the tentative spherical surface. The resulting aspheric surface is tangent to the tentative surface at the achromatic zone.

The system was then analyzed for residual chromatic aberration by tracing rays at the extreme wave-lengths through the system. The longitudinal displacements, from the focus of the rays of the most important wave-length, are shown graphically as a function of height of incidence in Fig. 1. This should be compared with a similar graph for a Schmidt system of somewhat smaller relative aperture ($f/0.83$) shown in Fig. 2. The improvement is obviously considerable in spite of the larger relative aperture of the hybrid system.

Although the improvement over previous systems is considerable, the balance of the residual sphero-chromatic aberration is not optimum. This optimum condition would be attained if the best focus (in the presence of residual aberration) for each of the extreme wave-lengths could be made to coincide with the perfect focus of the principal wave-length. If this is impossible the

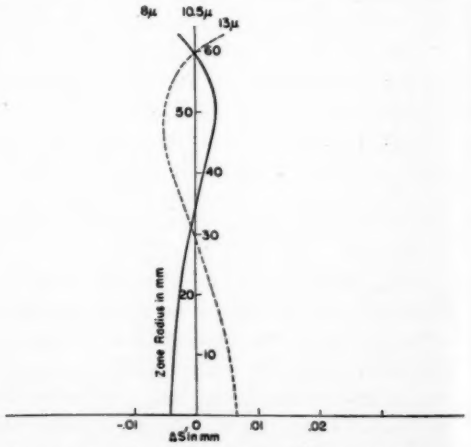


FIG. 3. Spherical aberration of Schmidt-Maksutov hybrid in which almost optimum balance of residual sphero-chromatic aberration has been achieved. Equivalent focal length = 89.66 mm.

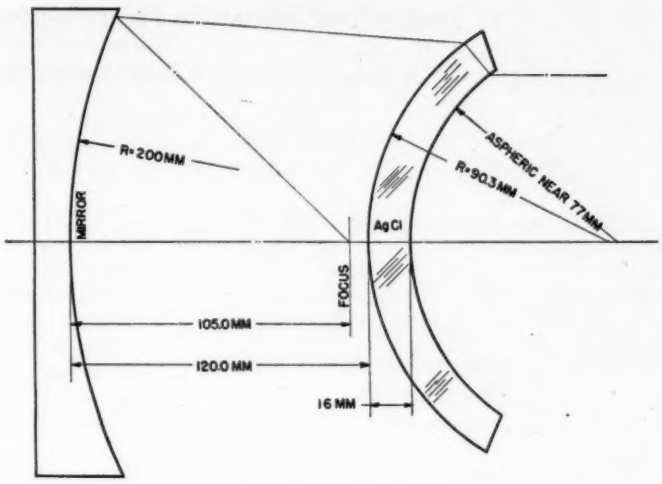


FIG. 4. Drawing of system whose aberrations are shown in Fig. 3.

next best adjustment would be such that the focal ranges[2] for the extreme wave-lengths overlap. Accordingly, further trials were made.

In succeeding designs the second radius of the corrector was so chosen that modest overcorrection of chromatic aberration was obtained at the selected zone. Then, when the aspheric surface had been substituted for the tentative first surface, correction of the type illustrated in Fig. 3 resulted. The residual optical path differences, at the focus of the principal wave-length, were less than one-tenth wave-length for the extreme wave-lengths. The coma of the system, as indicated b s maximum offense against the sine theorem + κ, 0.007, is greater than would usually be to ited, but was quite satisfactory for the purpo at hand. Figure 4 shows a drawing of this system.

The systems designed by the author were all for use in the infra-red region between 8μ and 13μ, 10.5μ being chosen as the principal wave-length. The principles involved in the reduction of the residual sphero-chromatic aberration should, however, apply equally well to systems for use in the visible or any other spectral range.

[2] A discussion of the "focal range" is given in Sir Richard Glazebrook, Ed., *Dictionary of Applied Physics* (Macmillan and Co., Ltd., London), Vol. IV, section on the optics of the microscope (written by A. E. Conrady), p. 221.

Installation of the Ohio Valley Section of the Optical Society of America

In the evening of April 27, 1948 at the Charles F. Kettering Foundation, Antioch College, Yellow Springs, Ohio, the Ohio Valley Section of the Optical Society of America was officially installed by Dr. Stanley S. Ballard, Secretary for Local Sections of the National Society, who subsequently addressed the group on the subject "Spectroscopy and Radiometry at Bikini."

This new local section is an outgrowth of the Ohio Valley Spectrographic Society, which was founded in 1944 in order to serve the technical interests of industrial spectroscopists and other optical people living in the area usually characterized as the Ohio Valley, which includes such industrial centers as Cincinnati, Dayton, and Middletown. The group has a membership of about fifty and meets approximately every six weeks throughout the year, usually distributing its meetings among Cincinnati, Middletown, Dayton, and Yellow Springs.

Officers of the new section are the following: Chairman, Dr. M. Eugene Merchant of the Cincinnati Milling Machine Company; Chairman-elect, J. F. Woodruff of the American Rolling Mill Company; Secretary-Treasurer, C. S. Mills, also of Armco; and Councilors: C. H. Aneshansley of the National Cash Register Company; Ivan T. Collier of the Moraine Products Division, General Motors Corporation; Wendell R. Koch of the Materials Laboratory, Wright Field; and Richard Olt of the Monsanto Chemical Company.

JOURNAL OF THE OPTICAL SOCIETY OF AMERICA VOLUME 38, NUMBER 8 AUGUST, 1948

Image Error Theory for Finite Aperture and Field

I. The Image of a Point; Geometry of the Wave Surface*,†

MAX HERZBERGER
Eastman Kodak Company, Rochester, New York

(Received March 4, 1948)

The author here suggests the replacement of the customary image error theory by a new one which permits the analysis of the symmetry qualities of the image by means of a few diagrams. Three new functions are introduced: the deformations function, D, the asymmetry function, A, and the spherical aberration function, S. The distance, τ, of the caustic from the diapoints is given by

$$\tau^2+S\tau+D\zeta^2=0,$$

and $D=S=0$ gives a system free from aberration. S, D, and A, are developed as functions of a special kind of variable, for which the author suggests $\eta-\eta_0$, $\zeta-\zeta_0$. Here, ζ and η are the direction cosines of the image ray with the optical axis and the perpendicular to it in the meridian plane; ζ_0 and η_0 are the corresponding angles for the principal ray of the bundle to be investigated.

Formulas are given for the development of these functions and for their connection with the characteristic function of the system.

IN various papers the author has stressed the point that the analysis of the image of an off-axis point in optical systems according to the conceptions taken from the Seidel theory is insufficient. He suggested that it be replaced by a new theory based upon the amount of symmetry or lack of it in the image.

This paper presents an attempt to develop the new theory in detail.

The rays coming from a point off the axis of an optical system form in the image space a bundle normal to a wave surface, having as the plane of symmetry the *meridian plane*, i.e., the plane containing both the object point and the axis.

The intersection points of the rays with the meridian plane are called the *diapoints* of the object points. It has been shown that the configuration of all the diapoints of a given image point are of importance for the image formation. If the diapoints form a curve, we say that the system is free from *deformation errors*, D. If the curve is a straight line, the image is said to be free of *symmetry errors*, A. In this case the only remaining error may be called *spherical aberration*, S, since it is comparable to the spherical aberration of an axis point. We shall in this paper derive the analytic conditions for freedom from these errors.

The aberration theory is generally developed in the following way: The characteristic function is developed into a series of suitable variables and in all further operations powers of an order higher than that of a given number are neglected. This procedure has the great disadvantage that the laws thus developed are only approximate, so that there might be no real system possible for which the geometric laws hold. As an example, we give the statement that the first-order (Gaussian) approximation gives a collinear relation between object and image rays and Sturm's theorem, according to which the rays of a normal bundle are, in first-order rays connecting two line elements, perpendicular to each other and to the rays at the primary and secondary focus.

The attempts made here consist in approximating a special characteristic function as usual by the section of its Taylor series, but thereafter calculating rigorously, without further approximation, the geometric qualities of the image formation given by the function chosen.

Thus, we consider only optically obtainable normal systems with a gradually increasing degree of complexity.

This method of course demands a careful scrutiny of the choice of variables, since the approximation must remain relatively simple. The best choice of variables would be the coordinates, y, z, of the diapoints. However,

* Communication No. 1162 from the Kodak Research Laboratories.

† This paper was read at the National Exhibition and Congress of Cartography and Optics, held in Florence, Italy, in October, 1947, and will be included in the published Proceedings of the Congress.

this cannot be done, for the simple reason that in a well-corrected optical system more than one image ray may have the same diapoint, i.e., the diapoints do not determine the image rays uniquely. The second-best choice, which we shall adopt here, is to take as coordinates the direction cosines, η and ζ, of the angles which the image ray makes with the y and z axis, respectively.

The aim of this paper is then to calculate in suitable coordinates the intersection points of a ray with the meridian plane and with a series of planes perpendicular to the axis (image planes). Furthermore, we want to calculate the caustic, and the conditions for freedom from *deformation errors*, D, *asymmetry errors*, A, and *spherical aberration*, S.

A NORMAL SYSTEM OF RAYS AND ITS SYMMETRY QUALITIES

We choose an origin in image space on the axis of the optical system and a Cartesian coordinate system such that the z axis is the optical axis and the yz plane is the meridian plane.

A ray is then given by the coordinates of one of its points for which we choose the diapoint (vector $\mathbf{a}$ with coordinates $x=0, y, z$), and its direction vector, $\mathbf{s}$, having the coordinates, ξ, η, and ζ, where

$$\xi^2+\eta^2+\zeta^2=1. \tag{1}$$

The rays coming from an object point form a two-dimensional manifold, that is, $\mathbf{a}$ and $\mathbf{s}$ can be considered as vector functions of two parameters, u and v. Later, we shall set $u=\eta$, $v=\zeta$; for the moment, we shall leave u and v arbitrary.

The well-known condition that the systems given by $\mathbf{a}(u, v)$ and $\mathbf{s}(u, v)$ form a normal system is given by the relation

$$\mathbf{a}_u\mathbf{s}_v=\mathbf{a}_v\mathbf{s}_u, \tag{2}$$

where $\mathbf{a}_u$ is an abbreviation for $d\mathbf{a}/du$, etc. Equation (2) is the integrability condition for the existence of a function, such that

$$\lambda_u=\mathbf{s}\mathbf{a}_u, \quad \lambda_v=\mathbf{s}\mathbf{a}_v, \tag{3}$$

and any integral λ of Formula (3) leads to a wave surface, $\mathbf{b}$, with

$$\mathbf{b}=\mathbf{a}+\lambda\mathbf{s}. \tag{4}$$

This is true because from

$$\mathbf{s}^2=1, \quad \mathbf{s}\mathbf{s}_u=\mathbf{s}\mathbf{s}_v=0, \tag{5}$$

and Eq. (4) it follows that:

$$\begin{aligned}\mathbf{b}_u\mathbf{s}&=\mathbf{a}_u\mathbf{s}-\lambda_u\mathbf{s}^2-\lambda\mathbf{s}\mathbf{s}_u=0,\\ \mathbf{b}_v\mathbf{s}&=\mathbf{a}_v\mathbf{s}-\lambda_v\mathbf{s}^2-\lambda\mathbf{s}\mathbf{s}_v=0.\end{aligned} \tag{6}$$

The caustic surface,

$$\mathbf{c}=\mathbf{a}+\kappa\mathbf{s}, \tag{7}$$

is a surface tangential to all the rays. It is found from the condition that the plane tangential to $\mathbf{c}$ contains the ray. This leads to the relation,

$$[\mathbf{c}_u, \mathbf{c}_v, \mathbf{s}]=0, \tag{8}$$

in which the brackets indicate the so-called triple product or determinant product of the three vectors.

Substituting for $\mathbf{c}$, according to Eq. (7), leads to the following quadratic equation for κ:

$$\begin{aligned}\kappa^2[\mathbf{s}_u, \mathbf{s}_v, \mathbf{s}]+\kappa([\mathbf{s}_u, \mathbf{a}_v, \mathbf{s}]\\ +[\mathbf{a}_u, \mathbf{s}_v, \mathbf{s}])+[\mathbf{a}_u, \mathbf{a}_v, \mathbf{s}]=0.\end{aligned} \tag{9}$$

The condition that the system is free from deformation errors is that the diapoints form a one-dimensional manifold.

This means that the vector product $(\mathbf{a}_u\times\mathbf{a}_v)$ vanishes, and with it

$$[\mathbf{a}_u\mathbf{a}_v\mathbf{s}]=0. \tag{10}$$

Equation (9) then has the solution $\kappa=0$, i.e., the diapoint curve forms one branch of the caustic, the other one being given by

$$\kappa=-\frac{[\mathbf{s}_u\mathbf{a}_v\mathbf{s}]+[\mathbf{a}_u\mathbf{s}_v\mathbf{s}]}{[\mathbf{s}_u\mathbf{s}_v\mathbf{s}]}. \tag{11}$$

Let us now choose, in the general case, $u=\eta$ and $v=\zeta$ as variables.

The vectors, $\mathbf{a}$ and $\mathbf{s}$, have then the coordinates $(0, y, z)$ and $[(1-u^2-v^2)^{\frac{1}{2}}, u, v]$, respectively.

Equation (2) gives

$$z_u=y_v, \tag{12}$$

which means that there exists a function, $\phi(u, v)$, such that

$$y=\phi_u, \quad z=\phi_v. \tag{13}$$

We find, as coordinates for the derived vectors,

$\mathbf{a}_u$, $\mathbf{a}_v$, $\mathbf{s}_u$, $\mathbf{s}_v$:

$$s_u:\left(\frac{-u}{(1-u^2-v^2)^{\frac{1}{2}}},\,1,\,0\right)\quad a_u:(0,\,\phi_{uu},\,\phi_{uv})$$
$$s_v:\left(\frac{-v}{(1-u^2-v^2)^{\frac{1}{2}}},\,0,\,1\right)\quad a_v:(0,\,\phi_{uv},\,\phi_{vv}). \tag{14}$$

The equation for the distance of the diapoint to the caustic becomes

$$\kappa^2+\kappa S+D(1-u^2-v^2)=0; \tag{15}$$

where

$$S=\frac{[\mathbf{a}_u\mathbf{s}_v\mathbf{s}]+[\mathbf{s}_u\mathbf{a}_v\mathbf{s}]}{[\mathbf{s}_u\mathbf{s}_v\mathbf{s}]}$$
$$=(1-u^2)\phi_{uu}-2uv\phi_{uv}+(1-v^2)\phi_{vv}, \tag{16}$$
$$D(1-u^2-v^2)=[\mathbf{a}_u\mathbf{a}_v\mathbf{s}]/[\mathbf{s}_u\mathbf{s}_v\mathbf{s}]$$
$$=(\phi_{uu}\phi_{vv}-\phi_{uv}^2)(1-u^2-v^2).$$

The diapoints are given by Formula (13), i.e.,

$$x=0,\quad y=\phi_u,\quad z=\phi_v. \tag{17}$$

The system is considered free from *deformation* if the diapoints form a curve

$$F(y,\,z)=0, \tag{18}$$

which leads to

$$D=\phi_{uu}\phi_{vv}-\phi_{uv}^2=0. \tag{19}$$

The system is free from *asymmetry* errors if the diapoints form a straight line:

$$\alpha y+\beta z+\gamma=0. \tag{20}$$

That means that ϕ is a solution of (19) and of the additional differential equation,

$$A=\phi_{uuu}\phi_{vvv}-\phi_{uuv}\phi_{uvv}=0. \tag{21}$$

If ϕ is given, we can therefore consider the function, D, as a measure of the deformation errors and A as a measure of the asymmetry errors. Doing this, we can speak of the deformation error or the asymmetry error of a single ray or of a bundle of rays, meaning the values of D and A for the coordinates of the rays.

If D equals zero, we can choose the coordinates, u, v, such that $\mathbf{a}$ is a function of u alone. In this case, the condition, (2), that the system is a normal system leads, because $\mathbf{a}_v=0$, to

$$\mathbf{a}_u\mathbf{s}_v=\mathbf{a}_v\mathbf{s}_u=0, \tag{22}$$

which shows that, integrated with respect to v, $\mathbf{a}_u\mathbf{s}$ is a function of u alone.

This means that all the rays through a point of the diapoint curve form a cone with the tangent of the diapoint curve as axis.

In the case of an undeformed image ($D=0$), the function given by the formula,

$$S=(1-u^2)\phi_{uu}-2uv\phi_{uv}+(1-v^2)\phi_{vv}, \tag{23}$$

determines the position of the caustic. If $S=D=0$, we find that ϕ must be a linear function of u and v, i.e., $y=\phi_u$ and $z=\phi_v$ are constant.

In this case, therefore, the system is free from all aberrations.

The diapoints are the intersection points of the rays with the meridian plane. There remains now the task of finding the intersection points with a plane $z=z_0$ perpendicular to the axis. From analytical geometry we find that

$$x=(z_0-\phi_v/v)(1-u^2-v^2)^{\frac{1}{2}},$$
$$y=(v\phi_u-u\phi_v/v)+z_0(u/v), \tag{24}$$
$$z=z_0.$$

For practical problems, the coordinates introduced here should be replaced by $\eta-\eta_0$ and $\zeta-\zeta_0$, where η_0 and ζ_0 are the direction cosines of the principal ray, since ζ_0, the cosine of the ray with the z-axis, is nearly always large. Calling these variables again u and v, we find that

$$D=\phi_{uu}\phi_{vv}-\phi_{uv}^2,$$
$$A=\phi_{uuu}\phi_{vvv}-\phi_{uuv}\phi_{uvv},$$
$$S=\phi_{uu}+\phi_{vv}-[\phi_{uu}(u+\eta_0)^2$$
$$+2\phi_{uv}(u+\eta_0)(v+\zeta_0)+\phi_{vv}(v+\zeta_0)^2]. \tag{25}$$

In the case of $D=0$, of course, it will be convenient to use the angle of the ray with the diapoint curve tangent as one of the variables.

Letters to the Editor

Measurement of Fluorescence Spectra

S. SAMBURSKY AND G. WOLFSOHN
Department of Physics, Hebrew University, Jerusalem
December 23, 1947

A TECHNIQUE has recently been described by Burdett and Jones for the photoelectric measurement of fluorescence spectra which is especially suited for spectra in the near ultraviolet region.[1] There is, however, one detail in the experimental arrangement which can easily lead to errors if the investigation is not restricted to a mere identification of fluorescent substances but aims at a complete analysis of their spectra. The cell containing the fluorescent solution is tilted in such a way that the rays forming the image of the fluorescent layer have to traverse the solution before reaching the monochromator. Hereby fluorescence light is absorbed by the solution, and this absorption might considerably affect the position as well as the intensity of the fluorescence maxima next to the absorption bands of the solute.

The results of the measurements given by the authors in Table II of their paper clearly demonstrate the danger involved in this arrangement. In the spectra of the benzene solutions both of anthracene and phenanthrene the bands of the shortest wave-length, i.e. those next to the absorption region, are completely suppressed in spite of the fact that the intensity of their peaks is approximately as strong as that of the next maximum, registered as the first one by the authors. In Table I the figures given by the authors are compared with those in a paper (apparently overlooked by them) published by the present writers on the fluorescence and absorption spectra of anthracene and phenanthrene in solutions.[2] The relative intensities of the maxima determined by the photographic method are given in the last column. The method used by the present writers, by which absorption of the fluorescence light is almost completely avoided, can briefly be stated as follows: A reduced image of the light source was formed immediately behind the window of the quartz cell. Fluorescence light emerging through the same window from this foremost layer of the solution was thrown on the slit of the spectrograph. By suitable adjustment the direction of geometrical reflection was avoided.

The obvious disadvantage of this method is the reduction of light intensity which is especially felt when small concentrations of the fluorescing matter are investigated. On the other hand, sensitivity of the optical system permitting, this arrangement is to be preferred, as the correct intensities as well as positions of the band maxima are thus reproduced even in those parts of the spectrum where they might well be distorted by absorption. It should not prove difficult to modify, in accordance with this method, the experimental arrangement of Burdett and Jones so as to adapt it to cases where a complete spectral analysis is required.

TABLE I. Wave-lengths and intensities of the fluorescent bands of anthracene and phenanthrene.

	Burdett and Jones	Sambursky and Wolfsohn	Relative intensities
Anthracene		382.8 mμ	1.6
	405 mμ	404.2	2.0
	428	428.0	1.0
	454	454.7	0.3
	487		
Phenanthrene			
		340.5	0.3
		348.0	1.9
	358.5	357.3	1.2
	367	365.6	2.2
	377	375.8	0.8
	385	385.2	1.0
	408	406.8	0.3

[1] R. A. Burdett and L. C. Jones, Jr., J. Opt. Soc. Am. **37**, 554 (1947).
[2] S. Sambursky and G. Wolfsohn, Trans. Faraday Soc. **36**, 427 (1940).

Ultraviolet Phase Microscopy

A. H. BENNETT, D. L. WOERNLEY, AND A. J. KAVANAGH
American Optical Company, Buffalo, New York
(May 28, 1948)

AN experimental ultraviolet phase microscope employing a 16 mm reflecting objective of the general type described by Brumberg[1] and Burch[2] has been designed and constructed. Other optical parts such as the condenser and diffraction plate were made from fused quartz.

In preliminary work with this optical system the source of radiation was an AH4 mercury vapor lamp (General

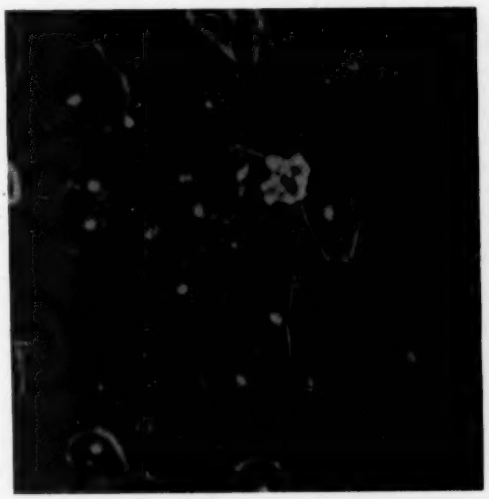


FIG. 1. Bright contrast with Hg 5461A monochromatic filter.

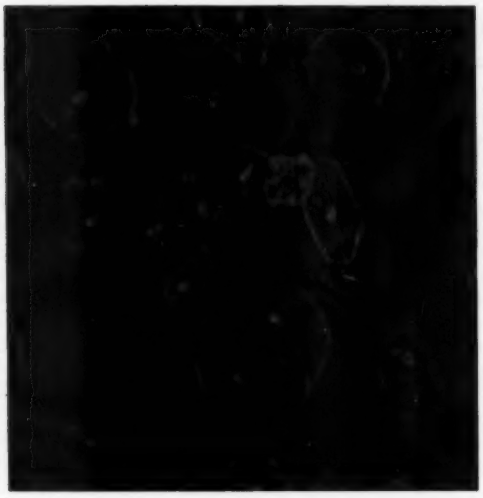


FIG. 2. Bright contrast with ultraviolet filter.

Electric Company) with its envelope removed. A Corning number 9863 filter provided a rather broad band of ultraviolet radiation extending approximately from 2300 to 4100 angstroms and another band commencing at about 6700A and extending into the near infra-red. Photographic film insensitive to the latter band was used. Since the reflecting type objective does not exhibit chromatic aberration and since it was sufficiently corrected with regard to other aberrations, it was feasible to use a broad band of ultraviolet radiation.

Figure 1 is a photograph of epithelial cells from the mouth with the 16 mm NA 0.30 reflecting objective, a bright contrast diffraction plate, and a Corning combination of numbers 512 and 351 filters to isolate the 5461A mercury line. The magnification was 32 times, and this picture was subsequently enlarged to 200 times. The photograph shown in Fig. 2 was taken under the same conditions as that in Fig. 1, except that the number 9863 ultraviolet filter was now employed. Both photographs were taken with exactly the same focal setting. At a magnification of 300 times, it is more apparent than is illustrated by Figs. 1 and 2 that the ultraviolet photographs are sharper than the visible, even though the focal setting was made with visible radiation. There also appears to be more detail in the ultraviolet photographs.

Of course, the phase retardation is different in the two examples shown, but this initial work has demonstrated that phase microscopy can be extended to the ultraviolet region of the spectrum. The results appear to be quite promising.

[1] E. M. Brumberg, Bull. Acad. Sci. U.S.S.R., Ser. Phys. **6**, 32 (1942).
[2] C. R. Burch, Proc. Phys. Soc. **59**, 41 (1947).

A Test on the Process of Photographing the Trajectory of a Rapidly Moving Distant Light Source

R. E. NUSBAUM AND J. S. BEGGS
North American Aviation, Inc., Los Angeles, California
June 28, 1948

THERE has been presented the problem of photographing, at night, the trajectory of a moving light with an object distance of about 20 miles. The proposed method is to use a fixed plate camera and make time exposures. As the light source moves across the field of view of the camera the image will produce a trace across the photographic plate. The effective exposure is determined by the intensity of the light source, the amount of light absorbed and scattered by the atmosphere, and the rate at which the image moves across the plate.

Calculations of the light intensity needed for a satisfactory trace are not reliable because the data on atmospheric absorption are rather vague. Since this application involves an expensive installation, the proposed method was tested under conditions similar to those to be experienced in practice.

In order to simplify the test, the movement of the light was simulated by rotating the camera about a vertical axis through a point near the first principal point of the lens. An Aerial K19B camera with a 12-inch, f/2.5 Aero Ektar lens and a Tri X Panchromatic photographic plate were used. The light source consisted of four tungsten filament lamps (photo-cell exciter lamps 7.5 A/T8/SC). These were mounted close together with the filament of each, a coil about $\frac{1}{16}$ inch in diameter and $\frac{5}{32}$ inch long, oriented with its axis perpendicular to the line of sight of the camera. A small polished metal reflector was mounted behind each lamp. The lamps were rated at 75 watts at 10 volts and were operated at 11 volts.

The first observation was made with the light source 8 feet above the ground at a point 20 miles from the camera. When the camera was rotated to move the image across the plate at the rate of 0.36 inch per second, the exposure was too low to produce a detectable trace. Although the observation was made on a relatively clear night the absorption of the atmosphere near the ground so greatly reduced the intensity of the light at the camera it could not be photographed with this sweep rate.

The light source was then mounted on an aircraft which was flown at an elevation of 11,000 feet and the camera was rotated in a direction opposite to the direction of flight. It was found that with the airplane 20 miles away and with a resultant motion of the airplane and camera producing a sweep rate of 0.4 inch per second the density of the photographic image of the trace near the center of the plate was 0.3. The plate was developed in DK60a developer.